

DOI: <https://doi.org/10.36910/6775-2524-0560-2023-53-39>

УДК 004.92

Черняшук Наталія Леонідівна, д.пед.н., професор

<https://orcid.org/0000-0002-3178-8377>

Семенюк Владислав Олегович, магістрант

Схабовський Максим Вадимович, магістрант

Токар Олександр Володимирович, магістрант

Оверчук Назар Богданович, магістрант

Луцький національний технічний університет, м. Луцьк, Україна

## РОЗПАРАЛЕЛЕННЯ ЗГОРТКОВИХ НЕЙРОННИХ МЕРЕЖ

Черняшук Н.Л., Семенюк В.О., Схабовський М.В., Токар О.В., Оверчук Н.Б. Розпаралелення згорткових нейронних мереж на основі графічних процесорів. В роботі класифіковано та досліджено розпаралелення алгоритмів навчання нейронних мереж із використанням технологій, що дозволяють використовувати графічні процесори для опрацювання великої кількості вхідних даних. Пришвидшення процесу аналізу біомедичних зображень згортковими нейронними мережами дозволяє в короткі терміни діагностувати будь-які відхилення у здоров'ї людини. Це дає можливість виявляти захворювання на ранніх стадіях та проводити лікування для запобігання їх поширення.

**Ключові слова:** згорткові нейронні мережі, графічний процесор, машинне навчання, штучний інтелект.

Chernyashuk N., Semenyuk V., Shabovskiy M., Tokar O., Overchuk N. Parallelization of convolutional neural networks based on graphics processors. The paper classified and investigated the parallelization of learning algorithms of neural networks using technologies that allow the use of graphics processors to process a large amount of input data. Acceleration of the process of analysis of biomedical images by convolutional neural networks makes it possible to diagnose any deviations in human health in a short period of time. This makes it possible to detect diseases in the early stages and carry out treatment to prevent their spread.

**Keywords:** convolutional neural networks, graphics processor, machine learning, artificial intelligence.

**Постановка наукової проблеми.** Загальні особливості функціонування згорткових нейронних мереж.

Нейронні мережі мають спеціальні шари, відомі як згорткові шари. Ці шари використовують фільтри для виявлення певних ознак у вхідних даних. Фільтри допомагають вилучити просторові особливості, такі як границі, форми та текстури, що є важливим для розпізнавання об'єктів на зображеннях. Після згорткового шару часто слідує підсумковий шар (пулінг), який використовується для зменшення просторового розміру репрезентації та зменшення кількості параметрів. Популярні методи пулінгу включають максимальне пулінг, де вибирається найбільше значення у певному регіоні, або середнє пулінг, де обчислюється середнє значення. Після кількох згорткових і підсумкових шарів можуть слідувати повні з'єднані шари, які використовуються для класифікації об'єктів на основі вивчених ознак. До кожного згорткового шару та повного з'єданого шару застосовується функція активації, така як ReLU (Rectified Linear Unit), для введення нелінійності та вдосконалення можливостей моделі у вивченні складних залежностей. Ваги моделі оптимізуються під час процесу навчання, щоб мережа вивчала розпізнавати певні шаблони та ознаки у вхідних даних. Ці особливості роблять згорткові нейронні мережі ефективними для обробки зображень і допомагають у вирішенні завдань, таких як розпізнавання облич, класифікація об'єктів та сегментація зображень.

**Аналіз досліджень.** GNM містить один або кілька згорткових шарів, які об'єднані або повністю з'єднані, і використовує варіації багатшарових перцептронів. Згорткові шари використовують операцію згортки на вході та передають результат до наступного шару. 1 зображено типову структуру ЗНМ.

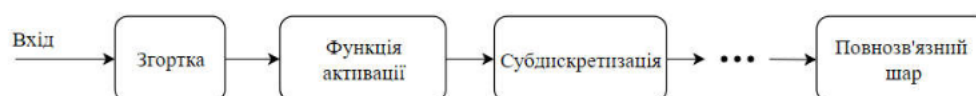


Рис. 1. Структура ЗНМ

Convolutional Neural Networks (CNNs) або загорткові нейронні мережі - це тип нейронних мереж, спеціально призначених для обробки структурованих сіток даних, таких як зображення. Ось деякі основні особливості їх функціонування:

*Загорткові шари (Convolutional Layers).* Основна ідея полягає в тому, щоб використовувати загорткові шари для локальної обробки вхідних даних, здійснюючи зсув і згортку (зменшення розміру) зображення, щоб виявити певні функції, такі як краї та текстури.

*Пулінгові шари (Pooling Layers).* Ці шари використовуються для зменшення розмірів зображення та виділення його важливих ознак. Зазвичай використовують максимальне або середнє знаходження значень у певному регіоні.

*Повнозв'язані шари (Fully Connected Layers).* Після декількох загорткових та пулінгових шарів використовуються повнозв'язані шари для класифікації або регресії. Вони об'єднують інформацію з попередніх шарів для прийняття рішення.

*Функції активації та нормалізація.* Застосування функцій активації (наприклад, ReLU) і методів нормалізації (наприклад, batch normalization) допомагає збільшити ефективність та стійкість мережі.

*Застосування до комп'ютерного зору.* CNN широко використовується для завдань комп'ютерного зору, таких як класифікація зображень, визначення об'єктів та розпізнавання обличчя.

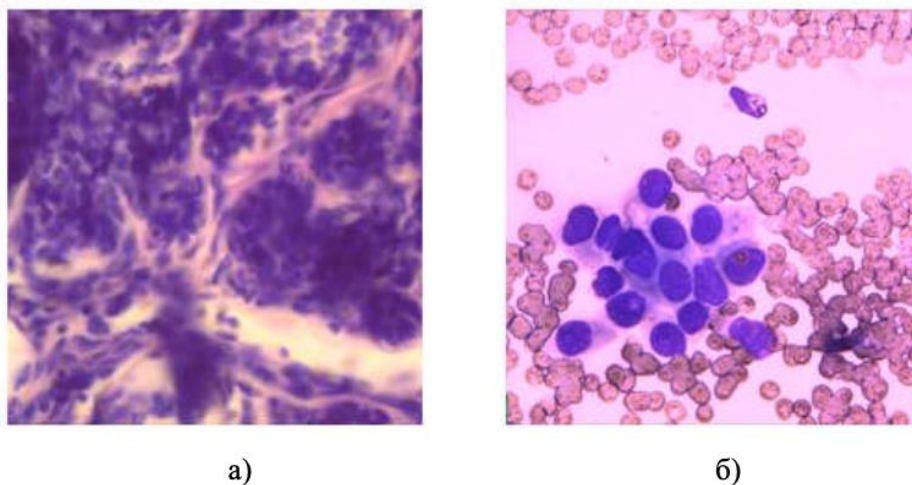
Ці характеристики роблять CNN дуже ефективними для обробки великих об'ємів зображень та виконання завдань, пов'язаних з комп'ютерним зором.

На рисунку 2 наведена послідовність кроків при класифікації зображень за допомогою ЗНМ.



Рис. 2. Алгоритм використання ЗНМ

На рисунку 3 показано відмінність гістологічного зображення від цитологічного. Гістологічне зображення складніше цитологічного за будовою, менш якісне та чітке.



а)

б)

Рис. 3. Фото:

а) гістологічні; б) цитологічні

**Мета роботи.** Метою даної роботи є аналіз алгоритмів навчання нейронних мереж та їхня паралелізація для виявлення найбільш ефективних способів реалізації на практиці, швидких і якісних програм для обробки великого обсягу інформації.

**Виклад основного матеріалу й обґрунтування отриманих результатів дослідження.**

Гістологічне дослідження – це метод детального дослідження організму людини з метою виявлення патологічного процесу на тканинному рівні. Цитологія – це наука, яка вивчає клітини, їх будову, функціонування, процеси розмноження, старіння та смерті [8].

Паралелізація є важливим аспектом оптимізації та прискорення роботи штучних нейронних мереж (ШНМ), особливо з урахуванням зростання розміру та складності моделей. Існує кілька методів паралелізації для штучних нейронних мереж:

*Методи паралелізації даних (Data Parallelism).* У цьому підході дані розбиваються на частини, і кожен обчислювальний пристрій (процесор чи GPU) обробляє свій фрагмент даних. Цей метод особливо ефективний для тренування глибоких нейронних мереж.

*Методи паралелізації моделі (Model Parallelism).* В цьому випадку сама модель розбивається на частини, і кожна частина обробляється окремим обчислювальним пристроєм. Цей метод корисний у випадках, коли модель занадто велика для одного пристрою.

*Мікро- та макро-паралелізація.* Мікропаралелізація використовується для розпаралелювання обчислень на рівні конкретних операцій, таких як перемноження матриць. Макропаралелізація, навпаки, включає розпаралелювання всієї моделі або її підсистем.

*Розподілене навчання (Distributed Training).* У цьому методі нейронні мережі навчаються на різних пристроях чи вузлах великих обчислювальних кластерів. Це може включати в себе як паралельне навчання моделі (model parallelism), так і паралелізацію даних (data parallelism).

*Асинхронна паралелізація.* Цей підхід передбачає виконання обчислень на пристроях асинхронно, тобто без чіткого синхронізованого обміну інформацією між пристроями під час тренування. Це може зменшити час очікування і збільшити швидкість навчання.

*Розпаралелювання на рівні операцій (Operation-level Parallelism).* Розпаралелювання може бути досягнуте на рівні окремих операцій у межах нейронної мережі, наприклад, використовуючи оптимізовані бібліотеки для виконання операцій паралельно на різних пристроях.

Ці методи можна комбінувати для досягнення оптимальної швидкодії та ефективності при навчанні та використанні штучних нейронних мереж. Використовуючи метод паралельної обробки під час фази навчання штучних нейронних мереж, можна досліджувати різні конфігурації мережі одночасно [9]. Таким чином, на рисунку 4 показано паралелізм тренувального сеансу, де різниця між фазами навчання представлена швидкістю навчання.

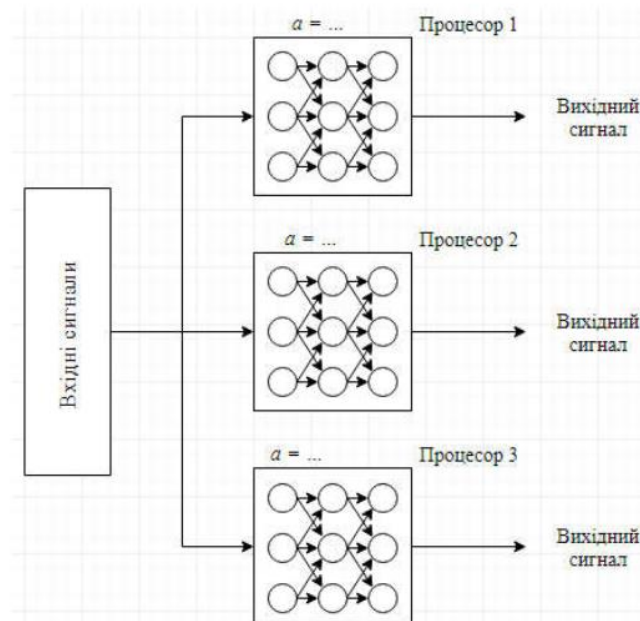


Рис. 4. Алгоритм фази навчання

При використанні розпаралелювання на рівні навчання вибірки навчаються одночасно на кількох різних навчальних вибірках. Це важливо, тому що часто вирішення великої задачі вимагає досить великої кількості векторів навчання в штучній нейронній мережі [11].

На рисунку 5 показана схема розпаралелювання навчальної множини.

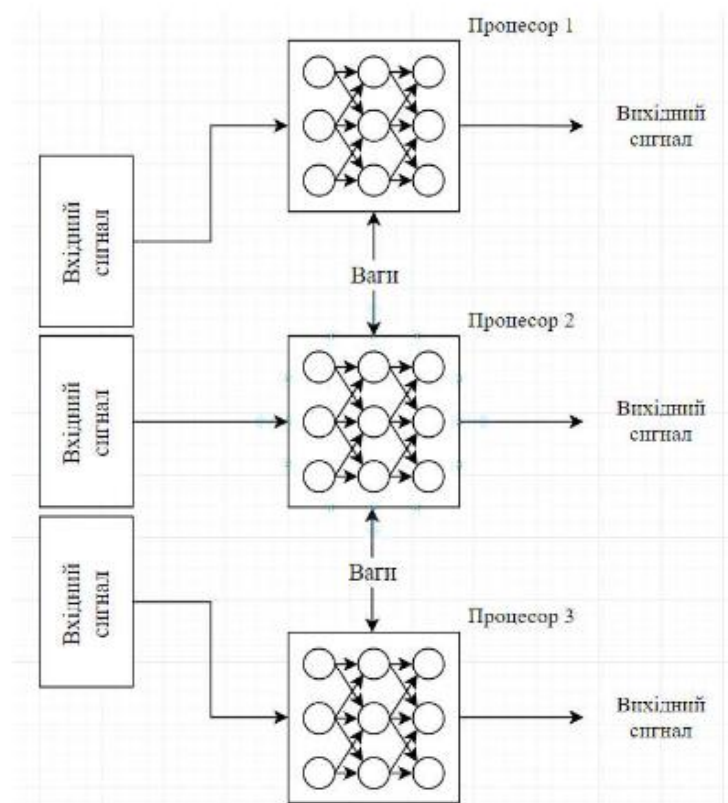


Рис. 5. Алгоритм навчальної вибірки

Це дозволяє перенести операцію на інший процесор. Той самий принцип можна використовувати в моделях, які не мають шарів штучних нейронних мереж, але вектори також проходять через процесори, як конвеєр. За еталон беруться найточніші вектори, що пройшли всі процесори. Коли вони знову проходять через мережеві процесори, процесори оновлюють шкали відповідно до цих стандартів. На рисунку 6 показана ця робоча схема.

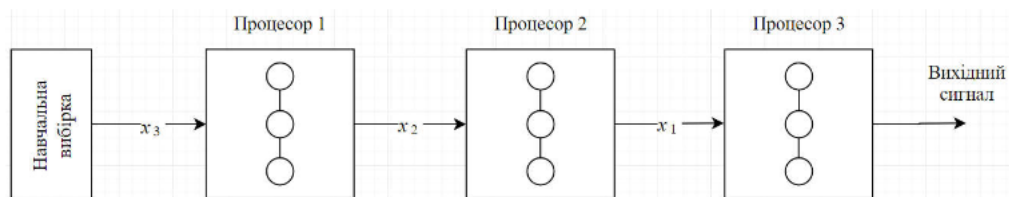


Рис. 6. Алгоритм паралелізації

При використанні методу розпаралелювання на рівні нейронів нейрон схожий за функціями на процесор, оскільки він також є елементом обробки. Паралельна обробка як процес на цьому рівні розділяє нейрони в межах одного рівня між процесорами, а також у паралельних обчисленнях. У цій системі обробки кожному процесору присвоюється нейрон або номер. Кожна з моделей штучної нейронної мережі має таку модель розпаралелювання і є одним із найпопулярніших методів. Схема роботи цього методу показана на малюнку 7.

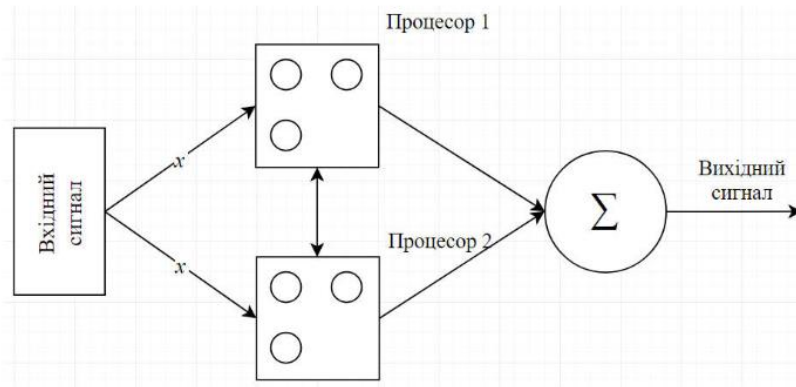


Рис. 7. Алгоритм розпаралелення [3]

Методом розпаралелювання є розпаралелювання на ваги. Схема типу розпаралелювання показана на рисунку 8.

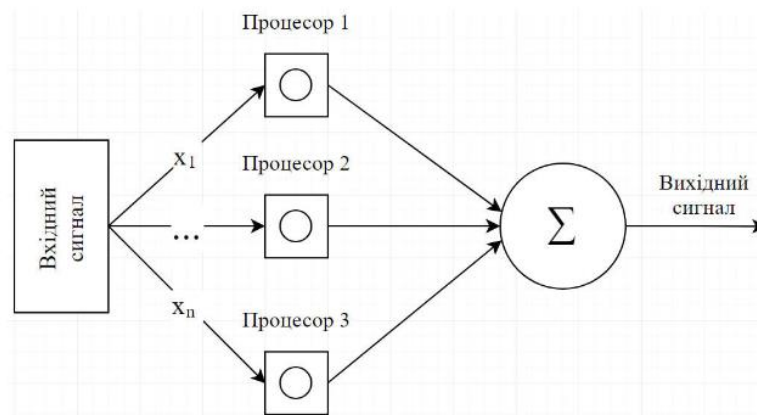


Рис. 8. Алгоритм розпаралелювання вагового рівня

Рівень вагових коефіцієнтів у методі розпаралелювання дрібнозернистий. Цей метод найчастіше використовується в апаратних реалізаціях. У цьому методі обчислення в межах одного нейрона розподіляються між кількома процесорами.

Щоб визначити ступінь успішності використання наведених методів розпаралелювання, потрібно створити систему відповідних характеристик.

**Висновки.** У роботі вирішується актуальна проблема розпаралелювання алгоритмів навчання згорткових нейронних мереж на базі GPU. Отримано наступні результати, проаналізовано методи розпаралелювання штучних нейронних мереж, переваги та недоліки використання кожного з них; досліджено сучасні технології та засоби розпаралелювання, в тому числі ті, що підтримують роботу з графічними процесорами, а саме: CUDA, OpenMP; проаналізовано алгоритми навчання згорткових нейронних мереж та їх порівняльну продуктивність при обробці великої кількості даних; розроблено схеми функціонування програмно-програмного засобу для паралельного навчання нейронних мереж за алгоритмом зворотного поширення з можливістю завдання користувачем основних параметрів роботи мережі; Було проведено ряд тестів, щоб підтвердити ефективність використання GPU для обробки великих обсягів даних за допомогою нейронних мереж.

Таким чином, використання графічних процесорів при роботі з нейронними мережами дозволяє отримати бажаний результат у багато разів швидше, ніж при використанні центральних процесорів.

#### Список бібліографічного опису

1. What is artificial neural network (ANN)? [Електронний ресурс] / – Режим доступу до ресурсу: <https://searchenterpriseai.techtarget.com/definition/neural-network>.
2. Розенблатт Ф. Принципи нейродинамики. Перейептрони и теория механизмов мозга / Ф. Розенблатт. – М.: Мир, 1965. – 175с.
3. Minsky M. L. Perceptrons / Minsky M. L. Papert S. A. – Cambridge, MA: MIT Press, 1969.

4. Werbos, P.J. Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences [Електронний ресурс] / – Режим доступу до ресурсу: <https://www.researchgate.net/publication/35657389>
5. Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Rumelhart, D.E; McClelland, James / Cambridge: MIT Press. ISBN 978-0-262-63110-5, 1986 – 567с.
6. What is big data? [Електронний ресурс] / – Режим доступу до ресурсу: <https://searchdatamanagement.techtarget.com/definition/big-data>
7. How Does the Brain Work? [Електронний ресурс] / Sandra Blakeslee // The New York Times – Nov. 11, 2003 – Режим доступу до ресурсу: <https://www.nytimes.com/2003/11/11/science/how-does-the-brain-work.html>.
8. Artificial Neural Network (ANN) in Machine Learning [Електронний ресурс] / –Режим доступу до ресурсу: <https://www.datasciencecentral.com/profiles/blogs/artificial-neural-network-ann-in-machine-learning>.
9. Yoon Kim Convolutional Neural Networks for Sentence Classification // Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP) – Doha, Qatar. – October 25-29, 2014. – P. 1746 – 1751.
10. Xiang Zhang Text Understanding from Scratch / Xiang Zhang, Yann LeCun // Computer Science Department, Courant Institute of Mathematical Sciences, New York University – New York, USA. – Apr 4, 2016.
11. Wen-tau Yih Semantic Parsing via Staged Query Graph Generation: Question Answering with Knowledge Base / Wen-tau Yih, Ming-Wei Chang, Xiaodong He, Jianfeng Gao // Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing – Beijing, China. – July 26-31, 2015. – P. 1321 – 1331.
12. Dasha Bogdanova Detecting Semantically Equivalent Questions in Online User Forums / Dasha Bogdanova, C'icero dos Santos, Luciano Barbosa† and Bianca Zadrozny // Proceedings of the 19th Conference on Computational Language Learning – Beijing, China. – July 30-31, 2015. – P. 123 – 131.
13. Ossama Abdel-Hamid Convolutional Neural Networks for Speech Recognition / Oссama Abdel-Hamid, Abdelrahman Mohamed, Hui Jiang, Li Deng, Gerald Penn, and Dong Yu // IEEE/ACM Transactions On Audio, Speech, And Language Processing, VOL. 22, No 10 – October 2014. – P. 1533 – 1545.
14. Recursive neural network [Електронний ресурс] / –Режим доступу до ресурсу: [https://en.wikipedia.org/wiki/Recursive\\_neural\\_network](https://en.wikipedia.org/wiki/Recursive_neural_network).
15. Natural Language Generation, Paraphrasing and Summarization of User Reviews with Recurrent Neural Networks [Електронний ресурс] / –Режим доступу до ресурсу: <http://www.meanotek.ru/files/TarasovDS292015-Dialogue.pdf>.
16. Siwei Lai Recurrent Convolutional Neural Networks for Text Classification / Siwei Lai, Liheng Xu, Kang Liu, Jun Zhao // Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence – China. – 2015. – P.2267– 2273.

#### References

1. What is artificial neural network (ANN)? [Electronic resource] / – Resource access mode: <https://searchenterpriseai.techtarget.com/definition/neural-network>.
2. Rosenblatt F. Principles of neurodynamics. Perieptrons and the theory of brain mechanisms / F. Rosenblatt. - M.: Mir, 1965. - 175p.
3. Minsky M. L. Perceptrons / Minsky M. L. Papert S. A. - Cambridge, MA: MIT Press, 1969.
4. Werbos, P.J. Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences [Electronic resource] / - Resource access mode: <https://www.researchgate.net/publication/35657389>
5. Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Rumelhart, D.E; McClelland, James / Cambridge: MIT Press. ISBN 978-0-262-63110-5, 1986 – 567с.
6. What is big data? [Electronic resource] / – Resource access mode: <https://searchdatamanagement.techtarget.com/definition/big-data>
7. How Does the Brain Work? [Electronic resource] / Sandra Blakeslee // The New York Times – Nov. 11, 2003 - Access mode: <https://www.nytimes.com/2003/11/11/science/how-does-the-brain-work.html>.
8. Artificial Neural Network (ANN) in Machine Learning [Electronic resource] / –Resource access mode: <https://www.datasciencecentral.com/profiles/blogs/artificial-neural-network-ann-in-machine-learning>.
9. Yoon Kim Convolutional Neural Networks for Sentence Classification // Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP) - Doha, Qatar. – October 25-29, 2014. – P. 1746 – 1751.
10. Xiang Zhang Text Understanding from Scratch / Xiang Zhang, Yann LeCun // Computer Science Department, Courant Institute of Mathematical Sciences, New York University - New York, USA. – Apr 4, 2016.
11. Wen-tau Yih Semantic Parsing via Staged Query Graph Generation: Question Answering with Knowledge Base / Wen-tau Yih, Ming-Wei Chang, Xiaodong He, Jianfeng Gao // Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing – Beijing, China. – July 26-31, 2015. – P. 1321 – 1331.
12. Dasha Bogdanova Detecting Semantically Equivalent Questions in Online User Forums / Dasha Bogdanova, C'icero dos Santos, Luciano Barbosa† and Bianca Zadrozny // Proceedings of the 19th Conference on Computational Language Learning – Beijing, China. – July 30-31, 2015. – P. 123 – 131.
13. Oссama Abdel-Hamid Convolutional Neural Networks for Speech Recognition / Oссama Abdel-Hamid, Abdelrahman Mohamed, Hui Jiang, Li Deng, Gerald Penn, and Dong Yu // IEEE/ACM Transactions On Audio, Speech, And Language Processing, VOL. 22, No. 10 – October 2014. – P. 1533 – 1545.
14. Recursive neural network [Electronic resource] / –Resource access mode: [https://en.wikipedia.org/wiki/Recursive\\_neural\\_network](https://en.wikipedia.org/wiki/Recursive_neural_network).
15. Natural Language Generation, Paraphrasing and Summarization of User Reviews with Recurrent Neural Networks [Electronic resource] / –Resource access mode: <http://www.meanotek.ru/files/TarasovDS292015-Dialogue.pdf>.
16. Siwei Lai Recurrent Convolutional Neural Networks for Text Classification / Siwei Lai, Liheng Xu, Kang Liu, Jun Zhao // Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence - China. – 2015. – P.2267– 2273.