

Д. М. Главчева, В. А. Яловега

Національний технічний університет «Харківський політехнічний інститут», Харків, Україна

## КАПСУЛЬНІ НЕЙРОННІ МЕРЕЖІ

**Предметом** вивчення є історія становлення та розвиток теорії нейронних мереж, сучасні підходи до проблем розпізнавання та класифікації зображень. Особлива увага приділяється якісному огляду капсульних та згорткових нейронних мереж, принципів їх роботи та визначення основних відмінностей. **Метою** роботи є аналіз сучасного стану досліджень нейронних мереж та можливих перспектив розвитку цієї галузі. **Завдання:** проаналізувати історичний розвиток теорії нейронних мереж. Провести порівняння між типами нейронних мереж, що базуються на концепції глибокого навчання: згортковими та капсульними. **Методом проведення дослідження** є аналіз сучасної літератури та основних тенденцій розвитку глибокого навчання. **Результатами** проведеного дослідження є виявлення значущих відкриттів, що вплинули на розвиток нейронних мереж. Функціонування нейронних мереж базується на роботі нервової системи біологічних організмів. Зокрема, це принцип активності біологічного нейрону, ансамблі нейронів, виявлення «простих клітин» у зоровій корі мозку. На даний момент найбільший розвиток мають нейронні мережі, що засновані на концепції глибокого навчання, яка дозволяє багатопаровим обчислювальним моделям вивчати дані з кількома рівнями абстракції. Згорткові мережі, що використовують цю концепцію досягли значних успіхів у розпізнаванні зображень, відео та аудіо. Рекурентні мережі виявилися кращі у аналізі тексту та мови. Згорткові нейронні мережі мають низку недоліків, на яких наголошено у роботі. Капсульні нейронні мережі є вдосконаленням концепції згорткових нейронних мереж. В їх основі покладено «капсули», які призначені для виявлення характеристик об'єкта. Капсули як група нейронів характеризуються вектором активації. Запропонований відомими ученими векторний підхід дозволяє врахувати поворот та трансляцію об'єктів. Капсульні нейронні мережі потребують значно меншу навчальну вибірку, ніж згорткові. У **висновках** роботи визначаються основні перспективи розвитку теорії нейронних мереж, а також можливий стрімкий розвиток неконтрольованого навчання нейронних мереж. Наголошується на важливості критичного аналізу проблем нейронних мереж як вирішального фактору їх майбутнього розвитку.

**Ключові слова:** глибоке навчання, капсульні нейронні мережі, згорткові нейронні мережі.

### Вступ

**Постановка задачі.** Протягом багатьох років людина намагається розгадати таємницю роботи свого мозку і створити високотехнологічні електронні пристрої, що здатні думати на рівні людського інтелекту. Людство ставить на меті автоматизування складних завдань, такі як розпізнавання та класифікація зображень, знаходження хворих кліток у організмі, відокремлення їх від здорових та інше. Для вирішення таких задач необхідно мати людський інтелект. Сучасні комп'ютери можуть швидко виконувати складні обчислення, які не під силу людині, проте задача розпізнавання та класифікації зображень є дуже складною для сучасних машин. Одним з перших результатів стала подія 1997 року, коли чемпіон світу з шахів Гаррі Каспаров зазнав поразки від комп'ютера IBM Deep Blue. Це викликало бурхливу хвилю росту досліджень у напрямку штучного інтелекту. Учених надихала видима простота мозку птахів та деяких комах у порівнянні зі складністю тих задач, які вони могли вирішити. Їх мозок демонстрував здатність керування польотом, вміння знаходити їжу, вирізняти хижаків та захищатися у небезпечних ситуаціях. Винахідників цікавило, чи зможуть комп'ютери з їх колосальними електронами ресурсами імітувати роботу мозку, не складнішого за мозок, наприклад, папуг [1]. Ідея проектування інтелектуальних обчислювальних пристроїв подібних до біологічних систем привела до створення теорії нейронних мереж, що побудована на основі організації та функціонування біологічних нейронних мереж.

**Метою роботи** є аналіз сучасного стану досліджень нейронних мереж та можливих перспектив розвитку цієї галузі.

### Результати досліджень

**1. Історичне становлення та розвиток теорії нейронних мереж.** Перша спроба формалізації нейронної мережі у математичну модель була розроблена у класичній роботі В. Маккалоха та В. Піттса [2]. Модель нейрона відповідала принципу активності біологічного нейрону – «все або нічого». Дослідники показали, що мережа, складена з великої кількості елементарних одиниць, здатна виконувати складні обчислення. Наступним етапом у розвитку нейронних мереж став вихід у світ в 1949 році книги Хебба (Hebb) *The Organization of Behavior* («Організація поведінки»). Зокрема, Хебб припустив, що, у міру того, як організм навчається різним завданням, зв'язки в мозку постійно змінюються і при цьому формуються ансамблі нейронів (*neuron assembly*). Постулат навчання Хебба стверджує, що ефективність змінного синапсу між двома нейронами підвищується при багаторазовій активації нейронів через даний синапс. У [3] була змодельована теорія нейронів Хебба. Результати моделювання чітко показали, що для повноти теорії Хебба слід додати корекцію синаптичних вагових коефіцієнтів. Учений Аттлі продемонстрував, що нейронні мережі із змінними синапсами можна навчити класифікації найпростіших растрових зображень. Він ввів поняття і активації нейрона, яке було формально проаналізовано Кайанелло (Caianello) в 1961 році [4].

У [5] стверджується, що адаптивна поведінка є не вродженою, а набутою, і за допомогою навчання можна поліпшити поведінку нейронної системи. У 1961 році була опублікована робота Мінського (Minsky), присвячена штучному інтелекту [6]. Вона

містила розділ, присвячений області, яка зараз називається теорією нейронних мереж.

Предметом окремого дослідження в контексті нейронних мереж є створення надійних мереж з нейронів. Це завдання було вирішене в 1956 році фон Нейманом за допомогою ідеї надмірності [7, 8]. У свій час Розенблатт (Rozenblatt) запропонував новий підхід до задачі розпізнавання образів, заснований на використанні перцептрона і нового методу навчання з учителем [9]. Головним досягненням цієї роботи була теорема збіжності перцептрона, перший доказ якої було отримано Розенблаттом у 1960 [10].

У 1982 Хопфілд використав функцію енергії для опису обчислень, виконуваних рекурентними мережами з симетричними синаптичними зв'язками [11]. У 1967 році Кован вів «сігмоїдальну» характеристику і гладку функцію активації для нейронів [12]. У 1986 році був розроблений алгоритм зворотного поширення помилки (back propagation algorithm) [13].

Згортова нейронна мережа (Convolutional neural network, CNN) була запропонована Яном Лекуном в 1988 році [14]. Її основна мета – ефективно розпізнавання зображень. Структура CNN не включає в себе зворотні зв'язки, тобто є односпрямованою та обов'язково багат шаровою. До сьогоднішніх днів CNN досить добре справляються із своїм призначенням, проте мають цілий ряд недоліків, що призвело до створення теорії капсульної нейронної мережі (Capsule neural network, CapsNet) видатним британським ученим Джеффрі Хінтоном (G. Hinton) [15].

**2. Основні недоліки згорткових нейронних мереж (ConvNet).** Глибоке навчання [16] дозволяє обчислювальним моделям, що складається з декількох шарів, вивчати дані з кількома рівнями абстракції. Ці методи в останній час значно допомогли у розпізнаванні мови, візуальних об'єктів та класифікації зображень. Глибоке навчання використовує алгоритм зворотного поширення помилки, щоб вказати, як система повинна змінити свої внутрішні параметри, які використовуються для обчислення в кожному шарі для кращого розв'язку поставленої задачі. Згорткові нейронні мережі (convolution) привели до прориву в обробці відео, зображень [17] та аудіо, тоді як рекурентні мережі показали успіхи при аналізі послідовних даних (текст і мова). Основний практичний успіх згорткових нейронних мереж - це розпізнавання обличчя. Продуктивність систем розпізнавання на основі згорткових мереж викликала використання їх більшістю великих технологічних компаній, включаючи Google, Facebook, Microsoft, IBM, Yahoo!, Twitter і Adobe, а також швидко зростаюче число стартапів для початку досліджень і розробок проектів і розгортання продуктів для розпізнавання на основі згорткової мережі [16]. Справді, можна сказати, що CNN показують гарні результати та справляються із своїми задачами на досить високому рівні, але вони мають суттєві недоліки, на які наголошує Хінтон [17].

Під час проектування логіки та архітектури згорткових нейронних мереж дослідники враховували особливості зорової кори, в якій були відкриті так звані прості клітини, що реагують на прямі лінії під різними кутами, і складні клітини, реакція яких пов'язана з

активацією певного набору простих клітин. Виходячи з цього, у CNN впроваджено чергування згорткових шарів і шарів pooling. Для навчання такої мережі найчастіше використовується метод зворотного поширення помилки, а функція активації – ReLU.

Однією з проблем згорткових нейронних мереж є наявність шару max pooling. Його основне призначення – зменшити розмірність матриць попереднього шару, не враховуючи деталі зображення, а звертаючи увагу тільки на ті дані, що дають найбільший внесок у розпізнаванні зображення. Якщо у зображенні наявні деякі характеристики як, наприклад, в обличчі людини є ніс, рот та очі, то незалежно від їх просторового розміщення, результат згорткової нейронної мереж залишиться незмінним. Мережа може помилково ідентифікувати обличчя, якщо характеристики (ніс, рот, очі) будуть розташовані у неправильному положенні відносно один одного. Шар max pooling можна не використовувати при проектуванні ConvNet, але все одно CNN не здатна враховувати трансляцію (translation) з поворотом (rotation). Такі серйозні проблеми призводять до того, що згорткову нейронну мережу необхідно навчати на великій кількості тренувальних даних, щоб вона мала здатність вирізняти одне й те ж саме зображення, повернуте на декілька градусів або трансльоване на відповідну величину. З іншого боку, наявність великої кількості змінних параметрів призводить до зменшення гнучкості CNN та їх використання тільки для дуже вузького класу задач. До параметрів, що можуть змінюватися можна віднести: кількість шарів, розмірність ядра згортки для кожного з шарів, кількість ядер для кожного з шарів, крок зсуву ядра при обробці шару, передавальну функцію нейронів. Всі ці параметри істотно впливають на результат та вибираються дослідниками емпірично.

**3. Капсульні нейронні мережі (CapsNet) як вдосконалення ConvNet.** Капсульні нейронні мережі спрямовані на усунення слабкості сучасних систем машинного навчання. Програмне забезпечення для розпізнавання зображень, яке сьогодні використовується, вимагає великої кількості прикладів зображень, щоб навчитися надійно розпізнавати об'єкти. Це пов'язано з тим, що відомі алгоритми майже не узагальнюють те, що вивчають.

Для врахування різних ракурсів, ієрархію простих та складних об'єктів у просторі, аналізу розташування елементів у зображенні була запропонована архітектура капсульних нейронних мереж. Такі мережі намагаються врахувати не тільки наявність необхідних характеристик у досліджуваному об'єкті, а й їх відносні координати. Це дозволяє більш чітко і точно класифікувати зображення та запроваджує інваріантність стосовно переміщення об'єкту в просторі.

Дослідження Хінтона [15] присвячено формалізації та ведення у нейронні мережі такого поняття як «капсули». Пропонується використовувати векторний підхід. У роботі автори визначають, що «капсула» - це група нейронів, вектор активації яких представляє параметри об'єкта або його частини. Вони призначені для відстеження різних частин об'єкта і їх положень у просторі. Використовується довжина ве-

ктора активності для представлення ймовірності існування характеристики (feature) та її орієнтації. Активні капсули на одному рівні роблять передбачення через матриці трансформації для параметрів капсул більш високого рівня. Коли кілька прогнозів співпадають, капсула більш високого рівня стає активною.

Капсули кодують ймовірність виявлення об'єкта як нормовану довжину вихідного вектора. Коли знайдений об'єкт рухається по зображенню, ймовірність його виявлення залишається незмінною (довжина вектора не змінюється), але змінюється орієнтація вектора. Нехай капсула визначила обличчя людини на зображенні і виводить 3D-вектор довжиною, наприклад, 0,96. При переміщенні частини тіла по зображенню вектор буде обертатися у своєму просторі, представляючи як змінюється положення обличчя, але його довжина залишиться фіксованою, тому що капсула знає, що вже виявила обличчя. Хінтон це називає еквіваріантністю діяльності нейронів: ймовірності виявлення об'єкта залишаються незмінними при зміні його положення.

Штучний нейрон (artificial neuron) описується такими характеристиками: скалярне зважування вхідних скалярів; сума зважених вхідних скалярів; скалярна нелінійність. З іншого боку, капсулу характеризують як векторне представлення нейронів: матричне множення вхідних векторів; скалярне зважування вхідних векторів; сума зважених векторів вхідних даних; векторна нелінійність.

Матричне множення вхідних векторів представляє собою операцію, в якій вхідні вектори перемножуються на матриці вагових коефіцієнтів. Такі матриці встановлюють зв'язок між характеристиками об'єкта нижнього рівня (очі, рот і ніс) і характеристикою вищого рівня (обличчя). Наприклад, може встановлюватися зв'язок між носом і обличчям: розмір носа менше розміру обличчя та знаходиться посередині. Отримані вектори представляють, де обличчя повинно знаходитись відповідно до виявлених положень очей, носу та рота. Скалярне зважування вхідних векторів відбувається за допомогою алгоритму «динамічної маршрутизації», описаного у [15]. Капсула нижчого рівня повинна «вирішити», до якої капсули вищого рівня відправляти свій результат детектування деякої характеристики досліджуваного об'єкта. Таке рішення вона приймає, після регулювання вагових коефіцієнтів, які будуть перемножуватися для формування результату, що отримує капсула вищого рівня як вхідні дані. Алгоритм «динамічної маршрутизації» між капсулами допомагає відрегулювати вагові коефіцієнти так, щоб капсулі вищого рівня передати більше «правильної» інформації, враховуючи дані, що були передані від інших капсул ниж-

чого рівня. Векторна нелінійність являє собою операцію яка нормує отриманий вектор так, щоб його довжина була не більше одиниці, при цьому напрямок вектора залишається незмінним:

$$\mathbf{v}_j = \frac{\|s_j\|^2 s_j}{1 + \|s_j\|^2 \|s_j\|},$$

де  $s_j$  – зважений вектор капсули.

У [15] показується, що дискриміновано навчена багатощарова система капсул є найпродуктивнішою в MNIST і результати значно кращі, ніж у згорткової нейромережі при розпізнаванні цифр.

## Висновки

Аналізуючи історію створення та становлення нейронних мереж, можна сказати, що людство намагається створити високоінтелектуальні обчислювальні пристрої, подібні до біологічних систем, адже навіть проста біологічна система здатна виконувати широкий спектр складних задач.

Теорія нейронних мереж є результатом плідної праці багатьох науковців. Ідеї щодо ансамблів нейронів, синаптичних вагових коефіцієнтів, теорема збіжності та алгоритм зворотного поширення помилки стали поштовхом до створення концепції глибокого навчання. У рамках цієї концепції найбільшого поширення отримали згорткові, рекурентні та капсульні нейронні мережі.

Інтерес до глибокого навчання істотно зріс після успіхів CNN. Цей тип нейронних мереж досяг прориву у розпізнаванні зображень, проте має певні недоліки, зокрема потребують велику вибірку для навчання та не враховуються взаємного розташування об'єктів, які були враховані при розробці ідеї капсульних нейронних мереж, що стала наступним кроком у теорії нейронних мереж і нині активно досліджується та вдосконалюється.

Ідея капсульних нейронних мереж показує ще перші, проте кращі результати при розпізнаванні цифр у зображеннях, а векторний підхід надає можливості аналізу характеристик зображень при їх повороті та трансляції. Ще однією перевагою капсульних нейронних мереж є зменшення кількості даних для тренування мережі. У майбутньому можна очікувати великий розвиток неконтрольованого навчання нейромереж, бо навчання людей і тварин в значній мірі неконтрольовано: ми отримуємо знання про навколишній світ, спостерігаючи його, а не повідомляючи завчасно властивості кожного об'єкта. З іншого боку, критичний аналіз проблем сучасних нейронних мереж стимулює пошук нових теорій та підходів, які будуть здатні якісно та точно вирішувати надскладні задачі розпізнавання та класифікації зображень.

## СПИСОК ЛІТЕРАТУРИ

1. Rashid T. Make your own neural network. – CreateSpace Independent Publishing Platform, 2016.
2. McCulloch W. S., Pitts W. A logical calculus of the ideas immanent in nervous activity/ W. S. McCulloch//The bulletin of mathematical biophysics. – 1943. – V. 5. – No. 4. – P. 115-133.
3. Rochester N., J.R. Holland, L.R. Haibt and W.L. Duda. Tests on a cell assembly theory of the action of the brain, using a large digital computer/ N. Rochester//IRE Transactions on Information Theory. – 1956. - V. IT-2. - P. 80-93.
4. Caianiello E.R. Outline of a theory of thought-processes and thinking machines //J.Theor.Biology. – 1961. – V. 1. - P. 204-235.
5. Anderson J.A., Pellionisz A. eds. Neurocomputing 2: Directions for Research, Cambridge, MA: MIT Press, 1990.
6. Minsky M.L. and S.A. Papert. Perceptrons, expanded edition, Cambridge, MA: MIT Press, 1988.

7. Bryson A.E., Jr. and Y.C. Ho. Applied Optimal Control, Blaisdell, 1969. (Second Edition, 1975, Hemisphere publishing, Washington, OC).
8. von Neumann J., Shannon C.E., McCarthy J. eds. Probabilistic logics and the synthesis of reliable organisms from unreliable components, in Automata Studies/ J. von Neumann/Princeton, NJ: Princeton University Press. -1956. - P. 43-98.
9. Rosenblatt F. The Perceptron: A probabilistic model for information storage and organization in the brain/ F. Rosenblatt//Psychological Review. - 1958, V. 65. - P. – 386-08.
10. Rosenblatt F. On the convergence of reinforcement procedures in simple perceptions/ F. Rosenblatt//Comell Aeronautical Laboratory Report. – 1960. – VG - 1196 - GA.
11. Hopfield J.J. Neural networks and physical systems with emergent collective computational abilities/ J. J. Hopfield//Proceedings of the National Academy of Sciences, USA. – 1982. - V. 79. - P. – 2554-2558.
12. Cowan J.D. A Mathematical Theory of Central Nervous Activity. Ph.D. Thesis, University of London, 1967.
13. Rumelhart O.E., G.E. Hinton and R.J. Williams. Learning representations of back-propagation errors/ O.E. Rumelhart//Nature (London). – 1986. - V. 323. - P. – 533-536.
14. LeCun Y. Backpropagation applied to handwritten zip code recognition // Neural computation. – 1989. – V. 1, No. 4. – P. 541-551.
15. Sabour S., Frosst N., Hinton G. E. Dynamic routing between capsules/ S. Sabour, N. Frosst, G. E. Hinton//Advances in Neural Information Processing Systems. – 2017. – P. 3856-3866.
16. LeCun Y., Bengio Y., Hinton G. Deep learning // Nature. – 2015. – V. 521, No 7553. – P. 436 – 444.
17. Tzeng E., Yang M., Ranzato, M. Deepface: closing the gap to human-level performance in face verification // Proc. Conference on Computer Vision and Pattern Recognition 1701–1708 (2014).

**Рецензент:** д-р техн. наук, проф. В. В. Вишневський,  
Державний університет телекомунікацій, Київ  
Received (Надійшла) 22.07.2018

Accepted for publication (Прийнята до друку) 22.09.2018

### Капсульные нейронные сети

Д. М. Главчева В. А. Яловега

**Предметом** изучения является история становления и развития теории нейронных сетей, современные подходы к проблемам распознавания и классификации изображений. Особое внимание уделяется качественному анализу капсульных и сверточных нейронных сетей, принципов их работы и определение основных различий. **Целью** работы является анализ современного состояния исследований нейронных сетей и возможных перспектив развития этой отрасли. **Задача:** проанализировать историческое развитие теории нейронных сетей. Провести сравнение между типами нейронных сетей, базирующихся на концепции глубокого обучения: сверточными и капсульными. **Методом** проведения исследования является анализ современной литературы и основных тенденций развития глубокого обучения. **Результатами** проведенного исследования является выявление значимых открытий, повлиявших на развитие нейронных сетей. Функционирование нейронных сетей базируется на работе нервной системы биологических организмов. В частности, это принцип активности биологического нейрона, ансамбли нейронов, выявление «простых клеток» в зрительной коре мозга. На данный момент наибольшее развитие имеют нейронные сети, основанные на концепции глубокого обучения, которая позволяет многослойным вычислительным моделям изучать данные с несколькими уровнями абстракции. Сверточные сети, используя эту концепцию достигли значительных успехов в распознавании изображений, видео и аудио. Рекуррентные сети проявили себя в анализе текста и языка. Сверточные нейронные сети имеют ряд недостатков, на которые указано в работе. Капсульные нейронные сети являются усовершенствованием концепции сверточных нейронных сетей. В их основе положен «капсулы», которые предназначены для выявления характеристик объекта. Капсулы как группа нейронов характеризуются вектором активации. Предложенный учеными векторный подход позволяет учесть поворот и трансляцию объектов. Капсульные нейронные сети требуют значительно меньшую обучающую выборку, чем сверточные. В **выводах** работы определяются основные перспективы развития теории нейронных сетей, а также возможность стремительного развития неконтролируемого обучения нейронных сетей. Отмечается важность критического анализа проблем нейронных сетей как решающего фактора их будущего развития.

**Ключевые слова:** глубокое обучение, капсульные нейронные сети, сверточные нейронные сети.

### Capsule neural networks

D. Hlavcheva V. Yaloveha

**The subject** of study is the history of the formation and development of the theory of neural networks, modern approaches to the problems of recognition and classification of images. Particular attention is paid to the qualitative review of capsular and convolutional neural networks, the principles of their work and the identification of the main differences. **The aim** of the work is to analyze the current state of neural network research and possible prospects for the development of this industry. **Objective:** to analyze the historical development of the theory of neural networks. Conduct a comparison between types of neural networks based on the concept of deep learning: convolutional and capsule. **The method of conducting** the research is an analysis of modern literature and the main trends of the development of deep learning. **The results** of the study are the discovery of significant openings that have influenced the development of neural networks. Functioning of neural networks is based on the work of the nervous system of biological organisms. In particular, this is the principle of the activity of the biological neuron, ensembles of neurons, the discovery of "simple cells" in the visual cortex of the brain. Currently, neural networks based on the concept of deep learning, which allows multi-layer computing models to study data with several levels of abstraction, are the most developed. Convolutional networks that use this concept have achieved significant success in recognizing images, videos and audio. Recurrent networks have appeared in the analysis of text and language. Convolutional Neural Networks have a number of shortcomings that are highlighted in the work. Capsule neural networks are an improvement in the concept of convolutional neural networks. They are based on "capsules", which are intended to detect the characteristics of the object. Capsules as a group of neurons are characterized by an activation vector. The vector approach proposed by researchers allows taking into account the rotation and translation of objects. Capsule neural networks require a much smaller training sample than convolutional. **The conclusions** of the work determine the main prospects for the development of the theory of neural networks, as well as the possible rapid development of uncontrolled training of neural networks. It is emphasized the importance of critical analysis of the problems of neural networks as a decisive factor for their future development.

**Keywords:** deep learning, capsule neural networks, convolutional neural networks.