

УДК 004.8

**Матківська Наталія Михайлівна**

*студентка*

*Навчально-наукового комплексу*

*«Інститут прикладного системного аналізу»*

*Національного технічного університету України*

*«Київський політехнічний інститут імені Ігоря Сікорського»*

## **БАГАТОРІВНЕВЕ НАВЧАННЯ**

***Анотація.** В даній статті розглянуто багаторівневе навчання, модель, його основні поняття та алгоритми, популярність та сфера використання.*

***Ключові слова:** нейронні мережі, програмне забезпечення, багаторівневе навчання.*

Багаторівневі нейронні мережі в даний час стають одним з найпопулярніших підходів до створення систем штучного інтелекту, таких як розпізнавання мови, обробка природної мови, комп'ютерний зір і т.п., а також популярним методом машинного навчання. Вони показують кращі результати в порівнянні з альтернативними методами в таких областях, як розпізнавання мови, обробка природної мови, комп'ютерний зір, медична інформатика та ін. Одна з причин успішного застосування багаторівневих нейронних мереж полягає в тому, що мережа автоматично виділяє з даних важливі ознаки, необхідні для вирішення завдання. В альтернативних алгоритмах машинного навчання ознаки повинні виділятися людьми, існує спеціалізований напрямок досліджень - інженерія ознак (feature engineering).

Однак при обробці великих обсягів даних нейронна мережа справляється з виділенням ознак набагато краще, ніж людина.

Модель штучних нейронних мереж була запропонована в 1943 році, а сам термін глибоке навчання (deep learning) став широко використовуватися тільки починаючи з 2006 року. До цього застосовувалися терміни завантаження багаторівневих мереж (loading deep networks) і навчання багаторівневої пам'яті (learning deep memories).

Зростання популярності багаторівневих нейронних мереж, що відбувається в останні кілька років, можна пояснити трьома факторами. По-перше, відбулося істотне збільшення продуктивності комп'ютерів, в тому числі прискорювачів обчислень GPU (Graphics Processing Unit), що дозволило виконувати навчання мереж значно швидше і з більш високою точністю. Раніше наявних обчислювальних потужностей не вистачало для навчання скільки-небудь складної мережі, придатної для вирішення практичних завдань. По-друге, був накопичений великий обсяг даних, який необхідний для навчання багаторівневих нейронних мереж. По-третє, розроблені методи навчання нейронних мереж, що дозволяють швидко і якісно їх навчати, що складаються зі ста і більше шарів, що раніше було неможливо через проблеми зникаючого градієнта і перенавчання. Поєднання трьох чинників призвело до істотного прогресу в навчанні мереж і їх практичному використанні, що дозволило багаторівневим нейронним мережам зайняти лідируючу позицію серед методів машинного навчання [1].

### **Штучні нейронні мережі та їх складові**

Штучні нейронні мережі були побудовані за принципом біологічних нейронних мереж, які представляють собою мережі нервових клітин, які виконують певні фізіологічні функції. Складовим елементом нейронних мереж є нейрони (представлені на рис. 1).

### Типичная структура нейрона

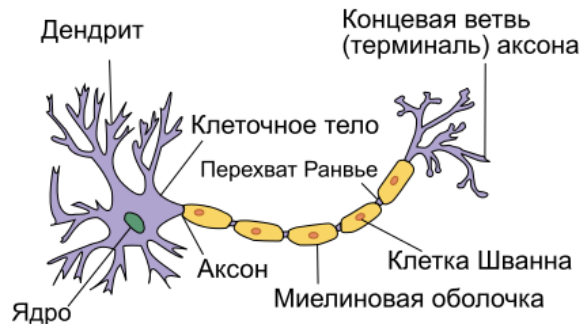


Рис. 1. Типова структура нейрона [2]

У нейрона є кілька функцій:

- Приймальня функція: синапси отримують інформацію;
- Інтегративна функція: на виході нейрона сигнал, який несе інформацію про суму сигналів в нейроні;
- Провідникова функція: по аксону проходить інформація до синапсів;
- Передаточна функція: імпульс, який досяг закінчення аксона, змушує медіатор передавати збудження наступному нейрону.

Синапсами називають зв'язку, по якій вихідні сигнали одних нейронів надходять на входи інших. Кожна зв'язка характеризується своєю вагою. Зв'язки з позитивним вагою називаються збудливими, а з негативним - гальмуючими. Вихід нейрона називається аксоном. У штучної нейронної мережі штучний нейрон – це деяка нелінійна функція, аргументом якої є лінійна комбінація всіх вхідних сигналів. Така функція називається активаційною. Потім результат активаційної функції посилається на вихід нейрона. Об'єднуючи такі нейрони з іншими, отримують штучну нейронну мережу [2].

## Перцептрон

Елементарний перцептрон будується на основі сенсорних даних на вході – S-елементів, асоціативних елементів – А-елементів, і реагуючих елементів на виході – R-елементів. Набір S-елементів, пов'язаний з А-елементом, утворює асоціацію, і елемент активується після досягнення певного числа сигналів від S-елементів. А-елемент передає ваговий сигнал на суматорний R-елемент, і залежно від того, чи перевищує суму певної порога, R-елемент видає результат роботи перцептрона (рис.2).

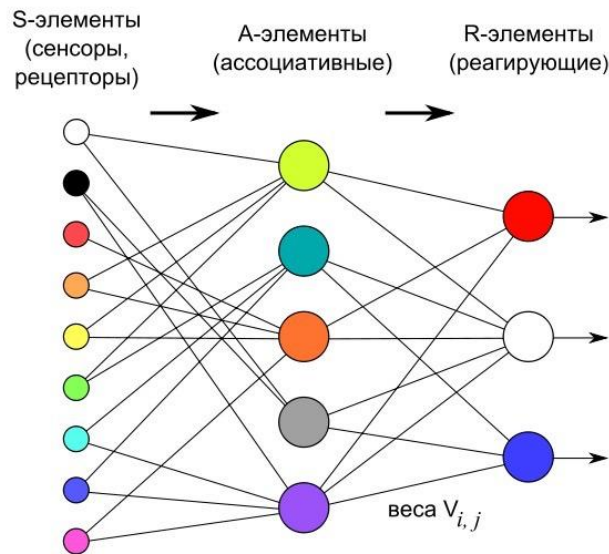


Рис. 2. Перцептрон [3]

Багатошаровий перцептрон будується з додатковими прихованими шарами А-елементів, розташованих між S-елементами і R-елементами. Принципіальна складність задач, що вирішуються багатошаровим перцептронним, є найвищою для класу перцептронів [3].

Навчання елементарного і багатошарового перцептрона полягає в зміні вагових коефіцієнтів зв'язків А - R. Перцептрон здатний працювати в режимі розпізнавання або узагальнення.

## **Багаторівневі нейронні мережі**

Багаторівневими нейронними мережами називаються такі мережі, в яких є кілька прихованих шарів. Оскільки кожен прихований шар обчислює нелінійне перетворення попереднього шару, багаторівнева мережа може мати значно більшу репрезентативну потужність (тобто може представляти значно складніші функції), ніж звичайна. При навчанні багаторівневої мережі важливо використовувати нелінійну функцію активації в кожному прихованому шарі. Це пов'язано з тим, що безліч шарів лінійних функцій самі вираховували б тільки лінійну функцію введення  $i$ , отже, не були б більш виразними, ніж при використанні тільки одного прихованого шару [4].

Головним достоїнством багаторівневих мереж є стисле представлення достатньо великої множини функцій. Можна показати, що існують функції, які  $k$ -шарова мережа може представити стисло, а  $(k-1)$  – шарова мережа не може цього зробити, якщо тільки вона не має експоненціально великої кількості елементів в прихованих шарах [5].

### **Доступність даних**

За допомогою методу, описаного вище, можна покладатися тільки на марковані дані для навчання. Однак помічених даних часто буває недостатньо, і, отже, для багатьох завдань важко отримати достатню кількість прикладів для відповідності параметрам складної моделі. Наприклад, з огляду на високий ступінь виразності багаторівневих мереж, навчання при невеликій кількості даних призведе до перенавчання.

### **Локальний оптимум**

Навчання малошарової мережі (з 1 прихованим шаром) з використанням контрольованого навчання зазвичай призводить до зближення параметрів з відповідними значеннями. Але при навчанні багаторівневої мережі, це працює набагато рідше. Зокрема, навчання нейронної мережі з

використанням навчання з учителем включає в себе вирішення проблеми з невивпуклою оптимізацією (наприклад, мінімізація помилки навчання в залежності від параметрів мережі  $W$ ). У багаторівневій мережі з'являється велика кількість локальних оптимумів, тому навчання з градієнтним спуском перестає працювати [6].

### **Градієнтна дифузія**

При використанні методу зворотного поширення помилки для обчислення похідних, градієнти, які поширюються від вихідного шару до більш ранніх шарів мережі, швидко зменшуються в міру збільшення глибини мережі. В результаті похідна від загальної вартості по відношенню до ваги в більш ранніх шарах дуже мала. Таким чином, при використанні градієнтного спуску ваги ранніх шарів повільно змінюються і більш ранні шари не можуть багато чому навчитися. Цю проблему часто називають "дифузією градієнтів" (diffusion of gradients).

**Висновки.** Багаторівневі нейронні мережі в даний час стають одним з найпопулярніших підходів до створення систем штучного інтелекту, таких як розпізнавання мови, комп'ютерний зір, і т.п.. Штучні нейронні мережі були побудовані за принципом біологічних нейронних мереж, які представляють собою мережі нервових клітин, які виконують певні фізіологічні функції, а складовим елементом нейронних мереж є нейрони. Головним достоїнством багаторівневих мереж є стисле представлення достатньо великої множини функцій.

### **Література**

1. Форсайт Д.А., Понс Ж. Компьютерное зрение. Современный подход. - М.: Вильямс, 2004. – 928 с.

2. Уоссермен Ф. Нейрокомпьютерная техника: Теория и практика. - М.: Мир, 1992. – 184 с.
3. Хайкин С. Нейронные сети: полный курс, 2-е издание. – М.: Вильямс, 2008. – 1103 с.
4. Флах П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных. – М.: ДМК Пресс, 2015. – 400 с.
5. Schmidhuber, J. Deep Learning in Neural Networks: An Overview / J. Schmidhuber // Neural Networks. — 2015. — Vol. 61. — P. 85–117.
6. Rosenblatt, F. Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms / F. Rosenblatt. — Buffalo, N.Y. : Cornell Aeronautical Laboratory, 1961. — xviii, 622 p. — (Cornell Aeronautical Laboratory; Report no. VG-1196- G-8).