

О.С. Безпалько, О.М. Польшакова

ПОШУК ОПТИМАЛЬНОЇ НЕЙРОННОЇ МЕРЕЖІ ДЛЯ ІНТЕЛЕКТУАЛЬНОГО ПОМІЧНИКА

Анотація: Інтелектуальні помічники спрямовані на спрощення роботи з клієнтною базою різних підприємств і організацій. Також, їх використання зумовлене автоматизацією процесу надання послуг, що значно підвищує швидкість роботи з користувачами сервісу та є економічно вигідним, так як зменшує кількість працюючого персоналу та, зв'язані з ним, витрати. У рамках даної роботи здійснено аналіз та порівняння різних типів нейронних мереж під час пошуків найкращої моделі для реалізації інтелектуального помічника. Визначені основні аспекти функціонування моделей, їхні переваги та недоліки.

Ключові слова: інтелектуальний помічник, нейронні мережі прямого поширення, зустрічного поширення, глибокі нейронні мережі, рекурентні нейронні мережі.

Опис проблеми

На теперішній час усе більше й більше знаходять застосування різноманітні типи нейронних мереж. Це й вирішення безлічі математичних задач, моделювання впливу різних факторів на об'єкт дослідження, розпізнавання зображень, робота з фотографіями, у кінці кінців створення штучного інтелекту робота. Одна з речей, яка вже широко використовується у світі, але ще не так популярна в Україні — це інтелектуальні помічники, що являють собою програми, які допомагають здійснити певні дії користувачу.

Наприклад: замовити квитки на літак, відслідковувати, коли ці квитки будуть у продажі. Ще, в якості прикладу, можна розглянути помічника, який допоможе знайти вам потрібний одяг, виходячи з досить абстрактного опису по типу: «Потрібен недорогий теплий одяг, який зараз у моді». Тому, тема даного дослідження є досить актуальною.

Проаналізуємо доступні технології та визначимо тип нейронної мережі, використання якої буде найбільш доцільною для створення інтелектуального помічника.

Огляд існуючих рішень

Почнемо з розгляду мереж прямого поширення, а саме перцептрона Розенблатта. Перцептрон, який побудований на основі трьох основних частин. Спочатку інформація надходить на, так звані, елементи-сенсори, де й

відбувається початкова обробка. Після визначення основних особливостей відбувається певне абстрагування за допомогою асоціативних елементів. У свою чергу, сукупність асоціативних елементів формує відповідь за допомогою реагуючих елементів. Весь описаний процес схематично зображений на рисунку 1.

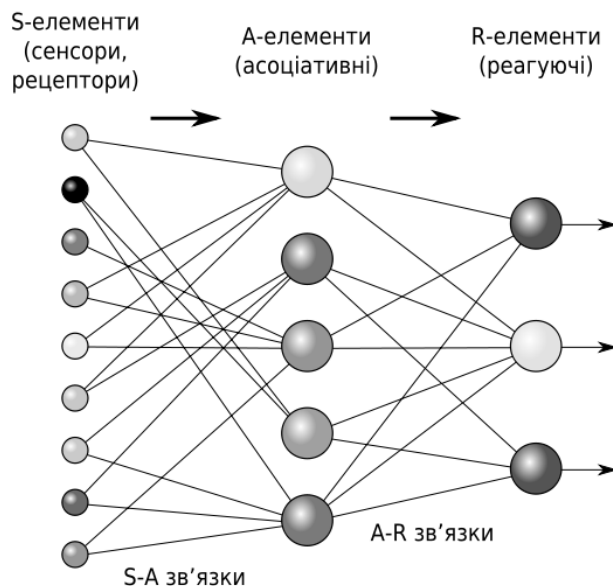


Рис. 1. Логічна схема перцептрону з трьома виходами.

Дана нейронна мережа є однією з найпростіших, але найчастіше вона застосовується для управління агентами та розпізнавання зображень. Виділимо переваги використання цієї технології:

- За рахунок не складної структури є досить простою у реалізації.
- Не потребує значних технічних та людських ресурсів, що робить її економічно вигідною.
- Високошвидкісна у роботі.
- Можливе вдосконалення моделі, шляхом використання навчання з вчителем, або методу зворотного поширення помилки.

До недоліків варто віднести:

- У наслідок простоти мережі не дозволяє побудову складних логічно-наслідкових ланцюгів.
- Набагато гірші показники точності при вирішенні задач, що включає великий обсяг інформації, яка не має прямих внутрішніх зв'язків.

Більш вдосконаленими нейронними мережами є багатошаровий перцептрон Розенблатта та нейронна мережа Ворда. Перша відрізняється тим, що, на відміну від звичайного перцептрона, містить не один шар A-елементів,

а декілька. Друга ж вирізняється тим, що нейрони прихованого шару поділяються на групи, які називаються блоками. Вони можуть мати різні функції обробки інформації, що дозволяє зробити аналіз об'єкта дослідження багатостороннім [1].

Перейдемо до розгляду рекурентних нейронних мереж. Основна особливість полягає у тому, що після першого навчання сигнал передається у зворотному напрямку. Типовим представником цього типу нейронних мереж є мережа Хопфілда, схема якої зображена на рисунку 2.

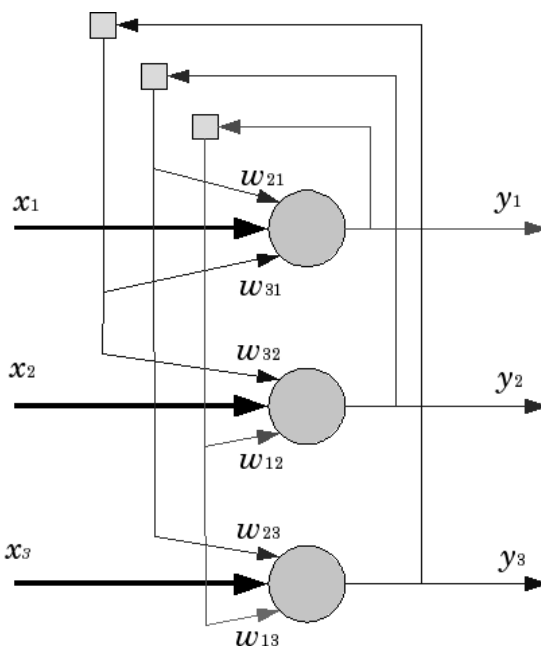


Рис. 2. Схема мережі Хопфілда з трьома нейронами.

Алгоритм, за яким функціонує дана мережа, сильно відрізняється у порівнянні зі стандартними алгоритмами навчання. Справа у тому, що на відміну від звичного послідовного корегування до оптимального стану з прораховуванням похибки, усі коефіцієнти матриці знаходяться за єдиною формулою, і система вже готова до роботи [2].

Мережа функціонує наступним чином: проходить перше визначення коефіцієнтів зв'язку між шарами нейронної мережі, досліджуваний об'єкт пропускається через систему, на виході нейронів обчислюється різниця між отриманим значенням та еталонним, відбувається зміна ваги коефіцієнтів поки мережа не набуде стану рівноваги. Вона може працювати у двох режимах: синхронний та асинхронний.

Якщо робота мережі моделюється на одному процесорі, то при синхронному режимі послідовно проглядаються нейрони, однак їх стани змінюються не-

залежно, і не змінюються до тих пір, поки не будуть пройдені всі нейрони мережі. Коли всі нейрони переглянуті, їх стани одночасно (тобто синхронно, звідси і назва) змінюються на нові. Таким чином, досягається моделювання паралельної роботи послідовним алгоритмом.

При реально паралельному моделюванні, цей режим фактично означає, що час передачі для кожного зв'язку між елементами однаковий, що призводить до паралельної роботи всіх зв'язків, які одночасно змінюють свої стани, ґрунтуючись тільки на попередньому моменті часу. Наявність таких синхронних тактів, які можна легко виділити, і призводить до розуміння синхронного режиму. При синхронному режимі можливо (хоча і не завжди спостерігається) нескінченне чергування двох станів з різною енергією — так званий динамічний атрактор. Тому синхронний режим практично для мережі Хопфілда не використовується, і розглядається лише як основа для розуміння більш складного асинхронного режиму.

Асинхронний режим роботи мережі функціонує інакше. Спочатку обчислюється значення першого нейрона, а всі послідовні за ним знаходяться у режимі очікування. Після закінчення обрахунків першого нейрона, починається визначення значення другого, враховуючи попереднє знайдене значення. І ця процедура продовжується до того моменту, поки не буде досягнуто останнього нейрона. Зрозуміло, що при такій режим роботи унеможливує виникнення динамічного атрактору.

Найголовнішою перевагою, яка спостерігається для цього типу нейронних мереж є можливість відновлення початкового стану об'єкта, що досліджується. Конкретно для реалізації інтелектуального помічника це може надати можливість достатньо чіткого визначення основного змісту повідомлення, навіть, якщо воно написано не чітко.

Серед недолік варто виділити:

- Порівняно невеликий об'єм пам'яті.
- Отримання стійкого стану систему, не гарантує отримання правильного результату.

Далі розглянемо нейронні мережі зустрічного поширення. На відміну від мереж зі зворотним поширенням помилки ці мережі призначені для початкового швидкого моделювання. Автор мережі зустрічного поширення помилки Р. Хехт-Нільсен вдало об'єднав в одній архітектурі переваги можливості узагальнення мережі Т. Кохонена [3] і простоту навчання вихідної зірки Д. Гроссберга [4], внаслідок чого мережа отримала властивості, яких немає ні в одній з них окремо. Вона належить до класу мереж, що навчаються без вчителя.

Мережа працює з двійковими векторами, що складаються з нулів і одиниць, або безперервними. В результаті навчання вхідні вектори асоціюються

з вихідними і, коли мережа навчена, подача вхідних образів призводить до отримання вихідних. Правильний вихід може бути отриманий і тоді, коли вхід є неповним, або дещо невірним. Гіперповерхня, що отримується в результаті функціонування мережі, внаслідок принципу безперервності, дає можливість здійснювати прогнозування. Природно, що всередині гіперпаралелепіеда навчальних образів прогнозування буде більш точним, а при вирішенні задачі екстраполяції помилка буде значно більше. Схема даної мережі зображена на рисунку 3.

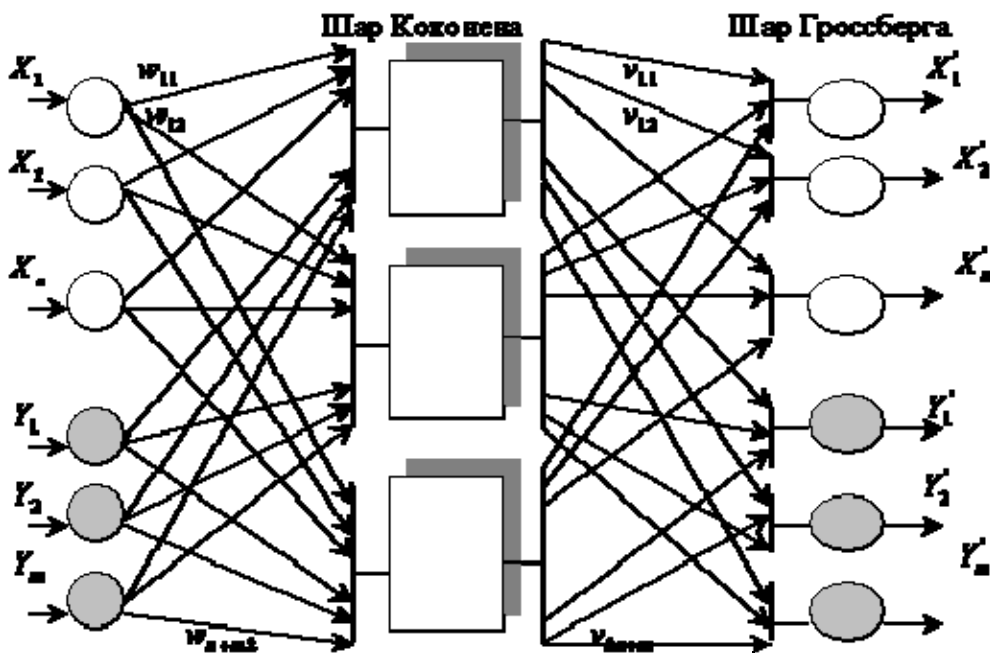


Рис. 3. Мережа зустрічного розповсюдження помилки.

Найчастіше дана система використовується спільно з мережею Хеммінга, варіацією якої є нейронна мережа Хопфілда, що була попередньо розглянута, в складі систем розпізнавання образів.

Недоліками вище розглянутої системи є — невідома кількість ітерацій, квадратичне зростання числа синапсів при збільшенні розмірності вхідного сигналу.

Перевага — простота.

Найскладнішими мережами у теперішній час є глибинні нейронні мережі. Розглянемо одного з представників глибинних нейронних мереж — згорткові нейронні мережі. У звичайному перцептроні, який представляє собою повнозв'язну нейронну мережу, кожен нейрон пов'язаний з усіма нейронами попереднього шару, і для кожного зв'язку існує свій персональний ваговий коефіцієнт. У згортковій нейронній мережі в операції згортки використовується лише обмежена матриця ваг невеликого розміру, яку «рухають» по всьому оброблюваному шару (на самому початку — без-

посередньо по вхідному зображенню), формуючи після кожного зсуву сигнал активації для нейрона наступного шару з аналогічною позицією. Тобто для різних нейронів вихідного шару використовується одна і та ж матриця ваг, яку також називають ядром згортки. Її інтерпретують як графічне кодування певної ознаки, наприклад, наявність похилої лінії під певним кутом. Тоді наступний шар, отриманий в результаті операції згортки такою матрицею ваг, показує наявність даної ознаки в шарі, який був оброблений, і її координати, формують так звану карту ознак. Природньо, коли в згортковій нейронній мережі набір ваг не один, а ціла гама, що кодує елементи зображення (наприклад лінії і дуги під різними кутами). При цьому такі ядра згортки не закладаються дослідником заздалегідь, а формуються самостійно шляхом навчання мережі класичним методом зворотного поширення помилки. Прокід кожним набором ваг формує свій власний примірник карти ознак, роблячи нейронну мережу багатоканальною (багато незалежних карт ознак на одному шарі). Також слід зазначити, що при переборі шару матрицею ваг її пересувають зазвичай не на повний крок (розмір цієї матриці), а на невелику відстань. Так, наприклад, при розмірності матриці ваг 5×5 її зрушують на один або два нейрона (пікселя) замість п'яти, щоб не «переступити» ознаку, яку шукають.

Операція субдискретизації, або агрегування виконує зменшення розмірності сформованих карт ознак. У даній архітектурі мережі вважається, що інформація про факт наявності шуканої ознаки важливіше точного знання його координат. Тому з кількох сусідніх нейронів карти ознак обирається максимальний і приймається за один нейрон ущільненої карти ознак меншої розмірності. За рахунок цієї операції, крім прискорення подальших обчислень, мережа стає більш інваріантною до масштабу вхідного зображення [5].

Розглянемо типову структуру згорткової нейронної мережі більш детально (рисунок 4). Мережа складається з великої кількості шарів. Після початкового шару (вхідного зображення) сигнал проходить серію згорткових шарів, в яких чергується власне згортка і агрегування. Чергування шарів дозволяє складати «карти ознак» з карт ознак. На кожному наступному шарі карта зменшується в розмірі, але збільшується кількість каналів. На практиці це означає здатність розпізнавання складних ієрархій ознак. Зазвичай після проходження декількох шарів карта ознак вироджується в вектор або навіть скаляр, але таких карт ознак стають сотні. На виході згорткових шарів мережі додатково встановлюють перцептрон, на вхід якого подаються кінцеві карти ознак.

Поговоримо про недоліки. Ця мережа найчастіше використовується для розпізнавання зображень. Тому використання її, як основи, для текстового консультанта є недоцільним. Детальне дослідження повідомлення принесе незначні покращення у точності результатів, але, при цьому, помітну затримку у часі ро-

боти самої системи. Враховуючи, що потік запитів може бути досить великим, це може призвести до значних та невиправданих затримок.

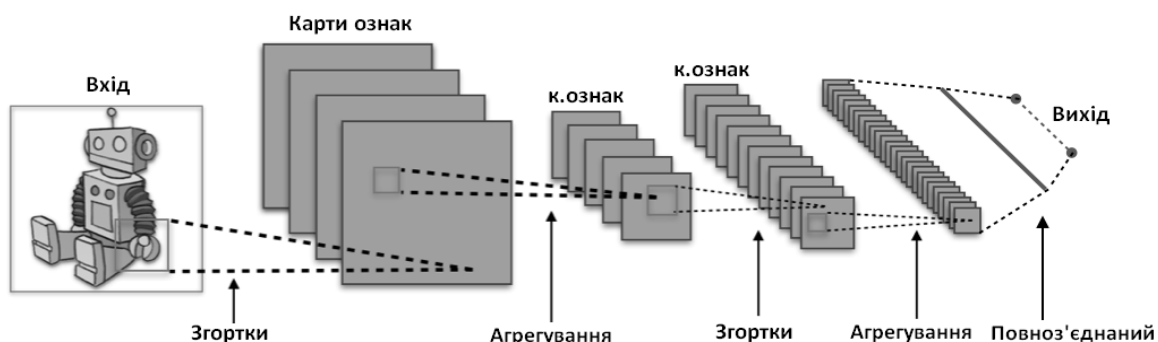


Рис. 4. Згорткова нейронна мережа.

Отже, якщо, наприклад, покупцю потрібно буде знайти якусь модель одягу за фото, то застосування загорткової нейронної мережі для інтелектуального помічника буде найбільш ефективним вирішенням такого типу задач. При цьому, будуть отримані найкращі показники точності разом із задовільною швидкістю.

Висновки

У рамках роботи була визначена актуальність досліджуваного питання, розглянуті різні типи нейронних мереж, проведений їх аналіз і, як результат, був сформований висновок, що при виборі оптимальної мережі для інтелектуального помічника потрібно враховувати особливості проекту для якого він буде використовуватися. Це й доступність фінансових та технічних ресурсів, обсяг вхідної інформації, формат, у якому ця інформація надається, вимоги до швидкодії та точності отриманих результатів та інше.

Список використаних джерел

1. B.V.Kryzhanovsky, B.M.Magomedov, A.L.Mikaelian. «A Domain model of neural network», Doklady Mathematics vol.71, pp. 310–314 (2005).
2. J. J. Hopfield. Learning algorithms and probability distributions in feed-forward and feed-back networks. — 1987.
3. Kohonen T. 1984. Self-organization and associative memory. Series in Information Sciences, vol. 8. Berlin: Springer Verlag.
4. Qrossberg S. 1969. Some networks that can learn, remember and reproduce any number of complicated space-time patterns. Journal of Mathematics and Mechanics, 19:53-91.
5. Уоссермен, Ф. Нейрокомпьютерная техника: Теория и практика. — М.: Мир, 1992. — 240 с. — ISBN 5-03-002115-9.