

РОЗДІЛ 5

ГЛИБИННІ НЕЙРОННІ МЕРЕЖІ

5.1 Глибинне навчання та глибинні нейромережі

Глибинне навчання (deep learning) – це розділ машинного навчання, спрямований на побудову ієрархічних моделей шляхом використання високорівневих масових абстракцій даних на основі глибинного графу з багатьма обробними шарами, які здійснюють лінійні або нелінійні перетворення, тобто це – прогресуючий високорівневий витяг ознак з сирих (необроблених) первісних вхідних даних.

У центрі глибинного навчання є глибинні нейронні мережі та методи їхньої побудови.

Глибинна нейронна мережа (ГНМ, deep neural network, DNN) – це різновид ШНМ, що має багато шарів обробки даних, яка перетворює вхідні дані у вихідні, ієрархічно виділяючи та агрегуючи ознаки, підвищуючи рівень абстракції даних в напрямку від входів до виходів.

Порівняно з мілкими НМ, ГНМ за рахунок збільшення кількості нейроелементів та зв'язків отримують більшу обчислювальну потужність і здатність моделювати більш складні залежності, а за рахунок спеціалізації шарів та високої ієрархічності обробки даних стають більш зручними для сприйняття та аналізу людиною. При цьому спеціалізація шарів обробки даних у ГНМ робить їх більш пристосованими до інтеграцію в мережеву модель апріорної інформації про предметну область. Проте, як правило, конкретні парадигми ГНМ мають більш обмежене застосування у конкретних задачах (наприклад, деякі архітектури можуть застосовуватися лише для розпізнавання зображень і не придатні для інших задач).

5.2 Згорткові нейромережі

Згорткові нейронні мережі (ЗНМ, convolutional neural network, CNN, ConvNet) – це глибинні ШНМ прямого поширення. Фактично вони є різновидом БНМ, адаптованої до обробки зображень.

Згорткові мережі (рис. 5.1) інспіровані організацією зорової кори тварин, де окремі нейрони кори реагують на стимули лише в обмеженій області зорового поля (*рецептивному полі*). Рецептивні поля різних нейронів частково перекриваються таким чином, що вони покривають усе зорове поле.

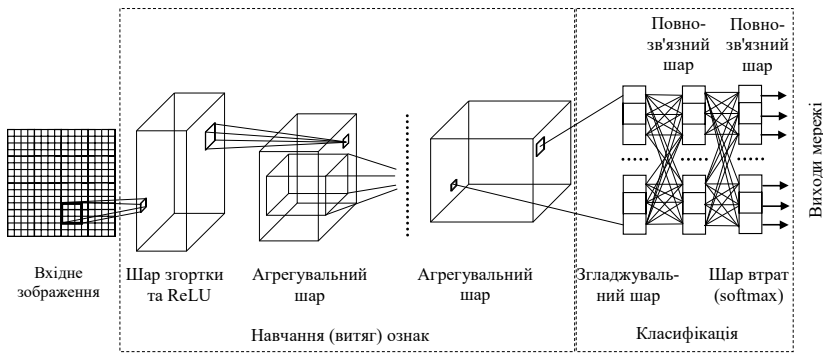


Рисунок 5.1 – Схема ЗНМ

ЗНМ складається з шарів входу та виходу, а також із декількох прихованих шарів. На відміну загальної парадигми БНМ, *приховані шари* ЗНМ зазвичай є спеціалізованими і складаються зі згорткових шарів, агрегувальних шарів, повноз'єднаних шарів та шарів нормалізації.

Згортковий шар (convolutional layer) застосовує до входу операцію згортки, передаючи результат до наступного шару:

$$\mathbf{z}^l = \mathbf{h}^{l-1} * \mathbf{W}^l,$$

де \mathbf{z}^l – постсинаптичний (передактиваційний) рівень збудження нейронів l -го шару НМ, \mathbf{h}^{l-1} – активаційний рівень нейронів $(l-1)$ -го шару НМ, \mathbf{W}^l – налагоджуваний параметр l -го шару НМ, $*$ – дискретний оператор згортки.

Згортка імітує реакцію окремого нейрона на зоровий стимул. Кожен згортковий нейрон обробляє дані лише для свого рецептивного поля.

Параметри шару складаються з набору фільтрів для навчання (або ядер), які мають невеличке рецептивне поле, але простягаються на всю глибину вхідної ємності. Протягом прямого проходу кожен фільтр здійснює згортку за шириною та висотою вхідної ємності, обчислюючи скалярний добуток даних фільтру та входу, і формуючи двовимірну карту збудження цього фільтру. В результаті мережа навчається, які фільтри активуються, коли вона виявляє певний конкретний тип ознаки у певному просторовому положенні у вході. Складання карт збудження всіх фільтрів уздовж виміру глибини формує повну ємність виходу

згорткового шару. Таким чином, кожен запис в ємності виходу може також трактуватися як вихід нейрона, що реагує на невеличку область у вході, та має спільні параметри з нейронами тієї ж карти збудження.

Агрегувальний шар (pooling layer) ЗНМ призначений для локального або глобального агрегування та об'єднує виходи кластерів нейронів одного шару до одного нейрону наступного шару. Агрегування є різновидом нелінійного зниження дискретизації, ідея якого полягає у тому, що точне положення ознаки не так важливе, як її грубе положення відносно інших ознак. Агрегувальний шар слугує поступовому скороченню просторового розміру подання для зменшення кількості параметрів та об'єму обчислень у мережі, і відтак також для контролю перенавчання. В архітектурі ЗНМ є звичним періодично вставляти агрегувальний шар між послідовними згортковими шарами. Операція агрегування забезпечує ще один різновид інваріантності відносно паралельного перенесення. Агрегувальний шар діє незалежно на кожен зріз глибини входу, і зменшує його просторовий розмір.

Максимізаційне агрегування (max pooling) використовує максимальне значення з кожного з кластерів нейронів попереднього шару. Воно розділяє вхідне зображення на набір прямокутників без перекриттів, і для кожної такої підобласті виводить її максимум:

$$h_{x,y}^l = \max_{i,j} \{h_{(x+i)(y+j)}^{l-1}\}, i = 0, \dots, s, j = 0, \dots, s,$$

де $h_{x,y}^l$ – рівень активації нейрона з індексами (x, y) l -го шару НМ.

Агрегування областей інтересу (Region of Interest pooling, RoI pooling) – це різновид максимізаційного агрегування, в якому розмір виходу фіксовано, а прямокутник входу є параметром:

$$h_{x,y}^l = \frac{1}{S^2} \sum_{i=0}^s \sum_{j=0}^s h_{(x+i)(y+j)}^{l-1}.$$

Усереднювальне агрегування (average pooling) використовує усереднене значення з кожного з кластерів нейронів попереднього шару.

$$h_{x,y}^l = \frac{1}{S^2} \sum_{i=0}^s \sum_{j=0}^s h_{(x+i)(y+j)}^{l-1}.$$

L^2 -нормове агрегування (L^2 -norm pooling) базується на Евклідовій відстані:

$$h_{x,y}^l = \sqrt{\sum_{i=0}^s \sum_{j=0}^s (h_{(x+i)(y+j)}^{l-1})^2}.$$

Шар зрізаних лінійних вузлів (Rectified Linear Units layer, ReLU) застосовує ненасичувальну передавальну функцію *ReLU*. Він посилює нелінійні властивості функції ухвалення рішення і мережі в цілому, не зачіпаючи рецептивних полів згорткового шару. Для посилення нелінійності застосовуються й інші функції, наприклад, насичувальні гіперболічний тангенс та сигмоїдна функція.

Згладжувальний шар (flatten layer) згортає просторові розмірності входу в розмірність каналу.

Повноз'єднаний шар (fully connected layer) з'єднує кожен нейрон одного шару з кожним нейроном наступного шару і забезпечує високорівневі міркування. Такий шар є різновидом шару традиційної БНМ:

$$\mathbf{z}^l = W^l \mathbf{h}^{l-1}.$$

Шар втрат (loss layer) визначає, як навчання штрафує відхилення між передбаченими та справжніми мітками класів, і є, як правило, завершальним шаром. Для різних завдань у ньому можуть використовувати різні функції втрат. Нормовані експоненційні втрати (*softmax*) застосовуються для передбачення єдиного класу з K взаємно виключних класів. Сигмоїдні перехресно-ентропійні втрати застосовуються для передбачення K незалежних значень імовірності в проміжку $[0, 1]$. Евклідові втрати застосовуються для регресії до дійснозначних міток $(-\infty, +\infty)$.

Основними відміннями ЗНМ від БНМ є наявність у ЗНМ:

– просторової організації (тривимірної ємності нейронів): згорткові шари ЗНМ мають нейрони, впорядковані у трьох вимірах: ширина, висота та глибина; нейрони всередині шару з'єднані лише з невеликою областю попереднього шару, що називається рецептивним полем; для формування архітектури ЗНМ складають різні типи шарів, як локально-, так і повноз'єднані;

– *локальної з'єднаності* між нейронами сусідніх шарів для забезпечення просторової локальності: навчені «фільтри» виробляють найсильніший відгук до просторово локального вхідного образу, а складання багатьох таких шарів веде до нелінійних фільтрів, що стають все глобальнішими (тобто, чутливими до більшої області піксельного простору), так що мережа спочатку створює подання дрібних деталей входу, а потім з них збирає подання більших областей;

– *спільних параметрів (ваг)*: для кожного рецептивного поля шару використовується один і той же фільтр (банк ваг), що це зменшує обсяг необхідної пам'яті та поліпшує продуктивність ЗНМ, бо кожен фільтр повторюється на всьому зоровому полі, а повторні вузли використовують спільну параметризацію (вектор ваг та упередженості) й формують карту ознак, що означає, що всі нейрони в заданому згортковому шарі реагують на одну й ту ж саму ознаку в межах свого рецептивного поля, повторювання вузлів таким чином дозволяє ознакам бути виявленими незалежно від їхнього положення в зоровому полі, забезпечуючи таким чином властивість інваріантності відносно зсуву;

– *шарів об'єднання*, у яких карти ознак поділяються на прямокутні підрегіони, а ознаки кожного прямокутника незалежно зменшуються до одного значення, зазвичай, беручи середнє або максимальне значення.

Разом ці властивості дозволяють ЗНМ досягати кращого узагальнення на задачах бачення. Спільне використання ваг різко зменшує кількість вільних параметрів, яких вчиться мережа, знижуючи таким чином вимоги до пам'яті для роботи мережі та уможливаючи тренування більших, потужніших мереж.

Навчання ЗНМ здійснюють шляхом використання техніки зворотного поширення помилки.

5.3 Мережі довгої короткочасної пам'яті

Мережа довгої короткочасної пам'яті (ДКЧП, long short-term memory, LSTM) – це архітектура, що відноситься до парадигми рекурентних нейронних мереж (РНМ), але на відміну від традиційних РНМ, мережа ДКЧП добре підходить вирішення задач класифікації, обробки або передбачення часових рядів, коли між важливими подіями існують часові затримки невідомої тривалості. Відносна нечутливість до довжини прогалін дає