

ВИКОРИСТАННЯ НЕЙРОННОЇ МЕРЕЖІ ДЛЯ РОЗПІЗНАВАННЯ ДРУКОВАНИХ СИМВОЛІВ

На сьогоднішній день існує досить багато потужних програм по розпізнаванню символів, але слід зазначити, що здатність людини читати друкований текст низької якості дотепер перевершує здатності комп'ютера.

При роботі з друкованим текстом хорошої якості OCR (optical character recognition) розпізнає символи практично зі 100 відсотковою точністю (99,9%). Почерк або стиль написання тексту - робить розпізнавання рукописного тексту набагато більш складним завданням, ніж розпізнавання друкованого тексту або навіть тексту, який написаний від руки друкованими літерами.

Нейронна мережа (neural network) більш всього інваріантна до різних спотворених зображень – це згортовка нейронна мережа, що навчається за методом зворотного розповсюдження похибки (back propagation)[1].

Ідея згортальних нейронних мереж полягає в чергуванні згортальних шарів (C-layers) рис.1, субдискретизуючих шарів (S-layers) і наявності повноз'язних (F-layers) шарів на виході.

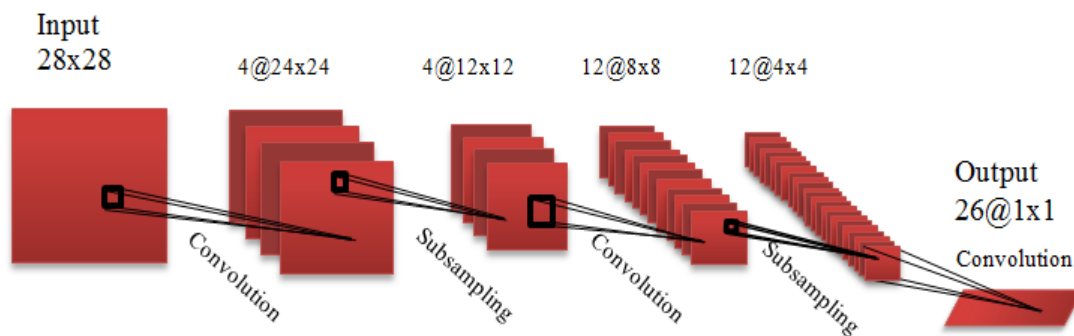


Рис. 1. Чергування згортальних шарів

Така архітектура містить в собі 3 основних парадигми:

– *Локальне сприйняття.* На вхід одного нейрона подається не все зображення, а лише деяка його область. Такий підхід дозволяє зберігати топологію зображення від шару до шару.

– *Спільні ваги.* Ця концепція ваг припускає, що для великої кількості зв'язків використовується дуже невеликий набір ваг. Тобто якщо на вході є зображення розмірами 28x28 пікселів, то кожен з нейронів наступного шару візьме на вхід тільки невелику ділянку цього зображення розміром, наприклад, 4x4. Важливо розуміти, що самих наборів ваг може бути багато, але кожен з них буде застосований до всього зображення. Такі набори часто називають ядрами (kernels).

– *Субдискретизація.* Суть субдискретизації і S-шарів полягає в зменшенні просторової розмірності зображення, яка потрібна для забезпечення інваріантності до масштабу.

Штучно введене обмеження на ваги покращує узагальнюючі властивості мережі (generalization), що в результаті позитивно позначається на здатності мережі знаходити інваріанти в зображенні і реагувати головним чином на них, не звертаючи уваги на інший шум.

Процес поширення сигналу в C-шарі (рис. 2):

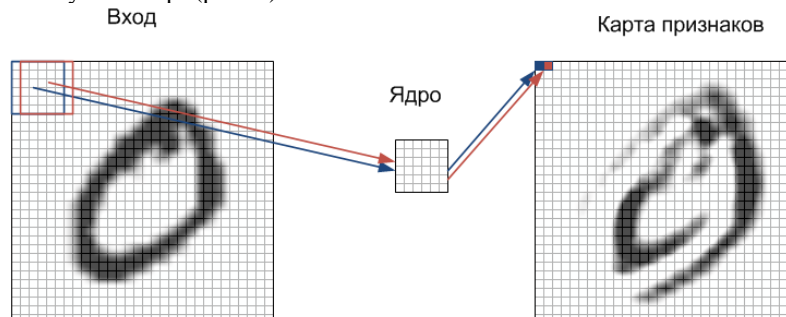


Рис. 2. Процес поширення сигналу

Кожен фрагмент зображення поелементно множиться на невелику матрицю ваг (ядро), результат підсумовується. Ця сума є пікселем вихідного зображення, яке називається картою ознак. Зважена сума входів ще пропускається через функцію активації:

$$OUT = \frac{1}{1 + \exp -\alpha Y}$$

де α - параметр нахилу сігмоїдальної функції активації. Змінюючи цей параметр, можна побудувати функції з різною крутизною.

Крім того, кількість ядер (наборів ваг) визначається розробником і залежить від того скільки ознак необхідно виділити. Ще одна особливість згортального шару в тому, що він зменшує зображення за рахунок крайових ефектів.

Чергування шарів дозволяє складати карти ознак. Зазвичай після проходження декількох шарів карта ознак вироджується в вектор або навіть скаляр, але таких карт ознак стає сотні. У такому вигляді вони подаються на один-два шари повно зв'язної мережі. Вихідний шар такої мережі може мати різні функції активації. У найпростішому випадку це може бути тангенціальна функція або радіальна базисна функція.

Для того щоб можна було почати навчання мережі потрібно визначитися з тим, як буде вимірюватися якість розпізнавання. Можна використовувати найпоширенішу в теорії нейронних мереж функцію середньоквадратичної помилки (СКО, MSE) [2]:

$$E^p = \frac{1}{2} D^p - O^p, W^2$$

де E^p - помилка розпізнавання для p -ої навчальної пари, D^p - бажаний вихід мережі, O^p - вихід мережі, що залежить від p -го входу і вагових коефіцієнтів W , куди входять ядра згортки, зміщення, вагові коефіцієнти S - та F - шарів. Завдання навчання так налаштувати ваги W , щоб вони для будь-якої навчальної пари (I^p, D^p) давали мінімальну помилку E^p . Щоб порахувати помилку для всієї навчальної вибірки просто береться середнє арифметичне по помилках для всіх навчальних пар. Таким чином E можна позначити як усереднену помилку.

Для мінімізації функції помилки E^p найефективнішими є градієнтні методи. Якщо розкласти в ряд Тейлора функцію помилки E^p , то отримаємо такий вираз:

$$E W = E W_c + W - W_c \frac{dE W_c}{dW} + \frac{1}{2} W - W_c^2 \frac{d^2 E W_c}{dW^2} + \dots$$

де E - функція помилки, W_c - деяке початкове значення ваги. Для знаходження екстремуму функції необхідно взяти її похідну і прирівняти нулю. Похідна функції помилки за вагами, відкинувши члени вище 2-го порядку буде:

$$\frac{dE W}{dW} = \frac{dE W_c}{dW} + W - W_c \frac{d^2 E W_c}{dW^2}$$

з цього виразу випливає, що ваги, при яких значення функції помилки буде мінімальним можна обчислити з наступного виразу:

$$W_{\min} = W_c - \frac{d^2 E W_c}{dW^2}^{-1} \frac{dE W_c}{dW}$$

Тобто оптимальна вага обчислюється як поточний мінус похідна функції помилки по вазі, поділена на другу похідну функції помилки. Для багатовимірного випадку (тобто для матриці ваг) все так же, тільки перша похідна перетворюється в градієнт (вектор приватних похідних), а друга похідна перетворюється в Гессіан (матрицю других часткових похідних). І тут можливі два варіанти. Якщо ми опустимо другу похідну, то отримаємо алгоритм найшвидшого градієнтного спуску. Зазвичай Гессіан замінюють чимось простішим, наприклад, метод Левенберга-Марквардта (ЛМ).

Представлені формули дозволяють легко обчислити похідну помилки по вазі, що знаходяться у вихідному шарі а обчислити помилку в прихованих шарах дозволяє широко відомий метод зворотного поширення помилки.

Список використаної літератури:

1. Y. LeCun and Y. Bengio: Convolutional Networks for Images, Speech, and Time-Series, in Arbib, M. A. (Eds), The Handbook of Brain Theory and Neural Networks, MIT Press, 1995.
2. Y. LeCun, L. Bottou, G. Orr and K. Muller: Efficient BackProp, in Orr, G. and Muller K. (Eds), Neural Networks: Tricks of the trade, Springer, 1998.