Проста згорткова нейронна мережа для розпізнавання рукописних цифр

В.В. Лукович

Міжнародний науково-навчальний центр інформаційних технологій та систем
40 просп. Академіка Глушкова, Київ 03680
E-mail: neuro@irtc.org.ua

Abstract

The aim of this paper was to investigate the performance of simple convolutional neural network. This network includes four layers. The first layer and the third layer are feature detectors whereas the second layer is a feature pooling layer. The last layer of the network act as linear classifier. The network was evaluated on the MNIST database. The on-line backpropagation algorithm was used to train this network. A character distortion method applied to increase the diversity of the training data. Experiments have shown this neural network can yield performance comparable to the state-of-the-art on handwritten digit recognition.

1. ВСТУП

Здатність багатошарових нейронних мереж, що навчаються за методом зворотного розповсюдження похибки (back propagation), реалізувати складні, багатовимірні та нелінійні відображення спонукає до використання таких мереж для розпізнавання зображень. У рамках традиційного підходу до розпізнавання образів блок попередньої обробки (препроцесор) виділяє релевантні ознаки зображень. Потім власне блок розпізнавання (класифікатор) категоризує отримані вектори ознак на класи. Але більш цікаво було б використати таку мережу, яка сприймала б на вході "сиру" інформацію та не потребувала б препроцесора. Після навчання кілька перших шарів такої мережі повинні виділяти релевантні ознаки зображень. На жаль, використання для цієї мети звичайних багатошарових персептронів з прямими зв'язками може бути пов'язане з певними проблемами.

По-перше, в практичних задачах розпізнавання зображення мають щонайменше кілька сотень елементів. Відповідно матриця зв'язків першого шару мережі з кількістю нейронів біля 100 може мати кілька десятків тисяч коефіцієнтів. В такому разі за недостатнього обсягу навчальної вибірки виникає небезпека того, що мережа "занадто добре" навчиться на нерелевантні варіації, присутні у зразках навчальної вибірки. Але головним недоліком неструктурованих мереж для розпізнавання зображень є відсутність "вбудованої" інваріантності до геометричних перетворень вхідного зображення та ігнорування його топології. Вхідні змінні можуть подаватися на мережу в довільному (але фіксованому) порядку без впливу на результат навчання. Проте відомо, що для зображень властива сильна локальна кореляція. Тому виявлення та комбінування локальних ознак перед розпізнаванням об'єктів часто підвищує ефективність розпізнавання.

2. ЗГОРТКОВІ НЕЙРОННІ МЕРЕЖІ

Архітектура згорткової нейронної мережі [1] поєднує виділення елементарних ознак зображення, формування більш складних ознак на вищих рівнях обробки та власне розпізнавання. Згорткова мережа в цілому навчається за єдиним алгоритмом (використовується певна модифікація широко розповсюдженого алгоритму зворотного розповсюдження похибки - back propagation). Ця мережа базується на таких принципах: нейрони мають локальні рецептивні поля; виходи нейронів – детекторів локальних ознак формують карти ознак зображення способом, подібним до математичної операції згортки (convolution); просторова роздільна здатність детекторів ознак зменшується на вищих рівнях мережі. Зображення в цілому не подається на входи нейронів першого шару мережі. Натомість ці нейрони послідовно "сканують" зображення своїми рецептивними полями з певним кроком. У процесі сканування вагові коефіцієнти нейронів не змінюються. У кожному положенні рецептивного поля вихідні сигнали нейронів першого шару формують зміст карти ознак.

Базовий модуль згорткової нейронної мережі складається з шару нейронів – детекторів ознак, вихід якого поступає на вхід шару нейронів, які виконують функцію групування ознак (feature pooling). Останні виконують функцію зменшення просторової роздільної здатності. Роль цього шару полягає у підвищенні стійкості представлення ознак до невеликих варіацій положення ознак на вході. Чергування вищезгаданих шарів дозволяє виділяти складніші ознаки з рецептивних полів, що зростають за розміром, при одночасному збільшенні стійкості до нерелевантних варіацій вхідних даних. Типова згорткова нейронна мережа складається з одного, двох або трьох з'єднаних послідовно модулів детектування/групування ознак, за якими йде модуль класифікації.

Як конкретні приклади згорткових нейронних мереж можна назвати LeNet5 [2], LeNet6 [3] та мережу, описану в роботі [4]. Входом мереж LeNet5 та LeNet6 є "сітківка" з розміром 32×32 піксели. В мережі LeNet5 перший шар детекторів ознак має 6 нейронів з квадратним рецептивним полем 5×5. Виходи цих нейронів формують першу карту ознак розмірністю 6×28×28. Перший шар групування ознак продукує карту ознак з розмірністю 6×14×14. Другий шар детекторів ознак формує карту ознак з розмірністю 16×10×10, нейрони цього шару мають рецептивні поля 5×5 та отримують вхідні дані з першої карти ознак. Другий шар групування ознак також зменшує вдвічі розмір карти ознак, так що на його виході друга карта ознак має розмірність 16×5×5. Наступний шар нейронів складається з 100 нейронів, підключених до

другої карти ознак. Останній шар нараховує 10 нейронів – по одному на кожен клас, що має розпізнаватися. LeNet6 має архітектуру, подібну до LeNet5, але кількість нейронів – детекторів ознак на кожному рівні мережі значно більша: 50 у першому, 50 у другому, передостанній шар мережі складається з 200 нейронів. LeNet5 має 60000 параметрів для навчання, LeNet6 – біля 315000.

Згорткова нейрона мережа, що описана у роботі [4], відрізняється тим, що не має шарів групування ознак. Нейрони – детектори ознак "сканують" вхідні поля з кроком, що дорівнює двом. Таким чином забезпечується зменшення просторової роздільної здатності карт ознак. Нейрони перших трьох шарів мають квадратні рецептивні поля 5×5. Нейрони другого та третього шару отримують на вхід сигнал з усіх площин вхідних карт ознак. Перший шар мережі нараховує 6 нейронів, другий – 60, третій – 100, останній – 10. Ця мережа має приблизно 180000 параметрів для навчання.

Метою даної роботи було дослідити спрощену певним чином архітектуру згорткової нейронної мережі. Розгляд загальної архітектури згорткових нейронних мереж дозволяє помітити наступне. Кілька перших шарів мережі є детекторами ознак. Перший шар відповідає за прості ознаки, наступні шари комбінують знайдені прості ознаки та формують більш складні ознаки. Також з зростанням номера шару відбувається поступове зменшення розмірності площин карт ознак, але одночасно зростає їх кількість. Зрештою, знаходиться шар нейронів, який вже фактично не виконує операції "згортки", бо розмірність рецептивних полів цих нейронів зрівнюється з розмірністю площин карти ознак на вході цього шару. Ці нейрони отримують всі дані з вищезазначеної карти ознак одночасно. Отже, можна зробити висновок, що цей шар вже належить до тієї частини нейронної мережі, яка відповідальна за розпізнавання зображення, поданого на вхід мережі. Відповідно всі попередні шари нейронної мережі виконують формування карт ознак.

Відомо, що 2-шаровий персептрон, що навчається за алгоритмом back propagation, може виконувати роль універсального класифікатора. Досить часто він і використовується в згорткових нейронних мережах. Для прикладу можна навести LeNet6 та мережу, описану в роботі [4]. Автором раніше досліджувалася згорткова нейронна мережа [5], в якій два останні шари були реалізовані на основі радіальних базисних функцій, які теж спроможні виконувати функцію універсального класифікатора. Варто уточнити, що універсальний класифікатор може розпізнавати класи, які мають складну структуру в просторі ознак, можуть створювати неопуклі та багатозв'язні області в просторі ознак.

Найпростішим класифікатором, що навчається, можна вважати лінійний класифікатор. Він розпізнає класи, які можливо розділити гіперплощинами у просторі ознак. Також відомо, що з збільшенням розмірності простору ознак ймовірність того, що класи можна розділити гіперплощинами, зростає. Повертаючись до згорткових нейронних мереж, підкреслимо, що формально розмірність вхідних даних на вході тієї частини мережі, що здійснює розпізнавання, є досить великою. Очевидно, вона дорівнює добутку кількості площин останньої карти ознак на розмірність площини. Наприклад, для мережі з роботи [4] це число дорівнює 1500 (60×5×5). Позаяк в задачі розпізнавання рукописних цифр кількість класів дорівнює всього 10, можна припустити, що лінійний класифікатор у складі згорткової нейронної мережі міг би непогано забезпечити розв'язок задачі розпізнавання. Необхідно зауважити, що дослідники, які створили архітектуру згорткових нейронних мереж, починали дослідження саме з такого варіанта, в якому класифікатор був лінійним. Це легко можна побачити, проаналізувавши архітектуру нейронної мережі LeNet1 [1]. Але вони потім перейшли до більш складних архітектур, не розкривши повністю потенціальні можливості простої архітектури.

2.1. АРХІТЕКТУРА НЕЙРОННОЇ МЕРЕЖІ

На рис. 1 зображена архітектура згорткової нейронної мережі, з якою виконувалися дослідження. Входом мережі є квадратна "сітківка" розміру 28×28. Перший шар нейронів – детекторів ознак нараховує 20 нейронів. Кожен з цих нейронів має квадратне рецептивне поле 8×8. Сітківка "сканується" цим рецептивним полем з кроком, що дорівнює одиниці. Отже, перша карта ознак FM_1 має

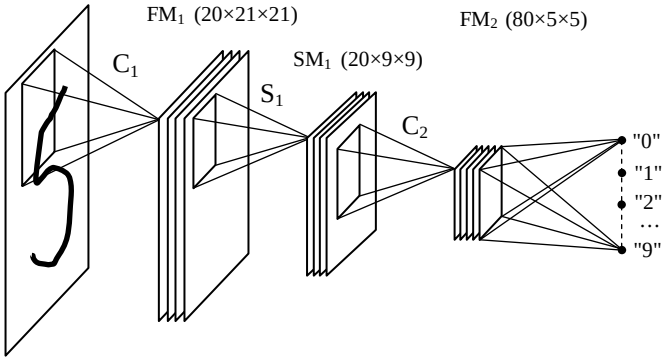

Рис.1. Архітектура простої згорткової нейронної мережі.

вигляд квадрата 21×21. Кількість площин в ній дорівнює 20, тобто кількості нейронів шару C_1. "Загрублена" перша карта ознак SM_1 отримується шляхом згортки кожної з площин карти ознак FM_1 з квадратною матрицею 5×5, яка є векторним добутком двох однакових 5-елементних векторів з елементами: {1,4,6,4,1}. Згортка виконується з кроком, що дорівнює двом, тому розмірність площин SM_1 становить 9×9. 80 нейронів – детекторів ознак третього шару мережі з рецептивними полями 5×5 оброблюють інформацію з карти ознак SM_1 та формують карту ознак FM_2 з розмірністю 80×5×5. Останній шар мережі утворюють 10 вихідних нейронів, кожен з них на вході має повну карту ознак FM_2. Загальна кількість параметрів для навчання мережі становить 61390.

Нейрони мережі мають неперервні вхідні та вихідні сигнали. Функція передачі нейронів має сигмоїдну нелінійність: $\sigma(S)=\frac{1}{1+\exp(-S)}$. Вихідні сигнали нейронів визначаються таким чином:

$$y=\sigma\left(\sum_k w_k x_k+\theta\right), \qquad (1)$$

де w_k – вагові коефіцієнти зв'язків нейрона, x_k – вхідні сигнали нейрона, θ – порог нейрона. Параметри w_k та θ змінюються при навчанні нейронної мережі. Індекс k в цій формулі є дещо умовним, для нейронів першого шару сумація виконується в межах двовимірного рецептивного

поля нейрона, для нейронів з наступних шарів додається ще третій вимір – номер площини карти ознак.

2.2. НАВЧАННЯ НЕЙРОННОЇ МЕРЕЖІ

Для навчання мережі був використаний метод зворотного розповсюдження помилки. За своєю суттю він є методом градієнтної мінімізації функції помилки мережі в просторі, утвореному параметрами нейронів мережі. Це означає, що навчання мережі здійснюється як ітеративний процес багаторазового обчислення поправок до параметрів нейронів, що складають мережу:

$$\Delta\xi = -\eta \cdot \frac{\partial E}{\partial \xi}, \tag{2}$$

та перерахунку параметрів нейронів

$$\xi' = \xi + \Delta\xi, \tag{3}$$

використовуючи знайдені поправки. Через ξ позначено один з вищезгаданих параметрів нейронів (w_k або θ). Функція помилки мережі E вибирається у вигляді

$$E = \sum_n e_n^2, \tag{4}$$

де

$$e_n = t_n - y_n \tag{5}$$

це різниця між t_n (цільовим, бажаним) та y_n (актуальним) значеннями n-го виходу мережі. Параметр навчання η є додатним числом, значно меншим за одиницю.

Для перерахунку параметрів нейронів мережі був використаний метод стохастичного градієнту (on-line). Цей метод передбачає перерахунок параметрів нейронів після кожного пред'явлення поточного зразка з навчальної вибірки. Процес навчання мережі було організовано у вигляді циклу. Кількість ітерацій цього цикла задавалася заздалегідь. Протягом кожної ітерації циклу вибиралися зразки з навчальної вибірки таким чином, щоб кожен з них за одну ітерацію один раз був поданий на вхід нейронної мережі. Початкове значення параметра навчання η встановлювалося рівним 0.05, після кожної ітерації η помножувалося на константу, меншу за одиницю. Таким чином, параметр навчання плавно зменшувався у процесі навчання мережі. З метою досягнення різноманітності послідовностей пред'явлення зразків з навчальної вибірки у різних ітераціях вищевказані послідовності організовувалися наступним чином. Стартова позиція першого зразка з навчальної вибірки формувалася як випадкове число з рівномірним розподілом (було використано стандартну функцію rand()). Також з масива послідовних простих чисел від 67 до 787 випадково вибиралося число – крок зміни позиції поточного зразка з навчальної вибірки. Позиція поточного зразка визначалася як сума попередньої позиції та кроку за модулем, що дорівнює розміру навчальної вибірки.

Перед навчанням задавалося певне порогове значення ε = 0.3 для помилки виходу мережі. Навчання мережі на поточному зразку здійснювалося не тільки за неправильної класифікації поточного зразка, але й при виконанні умови $\max_n |e_n| > \varepsilon$. Таким чином, навчання відбувалося не тільки на вхідних даних, що розпізнавалися неправильно, але й на таких, що розпізнавалися "не зовсім упевнено". Кількість зразків, для яких виконувався перерахунок параметрів мережі, підраховувалася протягом кожної ітерації. Якщо ця кількість ставала меншою за 4000, порог ε зменшувався шляхом множення на константу, меншу за одиницю.

З метою досягнення кращих показників навчання нейронної мережі використовувалося штучне збільшення обсягу навчальної вибірки шляхом деформації наявних зображень. Кожен зразок з навчальної вибірки перед тим, як бути поданим на вхід нейронної мережі, підлягав деформації. Параметри деформації генерувалися "на ходу", таким чином, навчальна вибірка віртуально отримала необмежену кількість зразків. Метод деформації близький до запропонованого в [6]. Він полягає у тому, що на зображення накладається квадратна сітка, вузли якої знаходяться у центрах пікселів зображення. Потім вершини квадрата незалежно один від одного зміщуються по обох кординатах, сітка відповідно деформується. Зміщення генеруються як рівномірно розподілені випадкові числа з нульовим середнім. Значення яскравості зображення вибираються в місці знаходження вузлів сітки після деформації, але проектуються на недеформовану сітку.

Аналогічно до роботи [3], для навчання були використані перших 50000 зображень цифр з навчальної вибірки. Решта 10000 зображень цифр використовувалися як валідаційна вибірка. Навчання мережі виконувалося протягом 150 ітерацій. Після кожної ітерації виконувалося розпізнавання валідаційної вибірки. Ітерація, на якій була досягнута мінімальна кількість помилок на валідаційній вибірці, фіксувалася. Стан мережі на цій ітерації вважався результатом процессу навчання і виконувалося розпізнавання тестової вибірки. Кращий результат розпізнавання тестової вибірки з серії експериментів наведений в таблиці 1 (0.38 відсотка помилок або 38 помилок з 10000 зображень, що належать до тестової вибірки).

Програмна реалізація була зроблена у вигляді консольної програми для операційних систем Windows2000/XP. Використано компілятор C/C++ з набору для розробки програмного забезпечення Microsoft Visual C++ Toolkit 2003. Для прискорення обчислень підпрограма знаходження сумарного вхідного сигналу нейрона була написана на асемблері з використанням команд з набору SSE.

3. БАЗА ДАНИХ

Для досліджень була використана база даних MNIST, яку склав Y. LeCun [2]. Вона є досить популярною для оцінювання якості алгоритмів розпізнавання та їх порівняння. В Інтернеті її можна знайти за адресою http://yann.lecun.com/exdb/mnist/. База даних містить 70000 зображень рукописних цифр від 0 до 9. 60000 зображень складають навчальну вибірку, контрольна вибірка має 10000 зображень цифр. Зображення мають розмір 28×28 пікселів. Рівень яскравості піксела кодується одним байтом. Навчальна та контрольна вибірки знаходяться в окремих файлах, також є ще два файли, які містять позначення класів.

4. РЕЗУЛЬТАТИ

В таблиці 1 наведено відсоток помилок розпізнавання контрольної вибірки з бази даних MNIST після навчання нейронної мережі. Для порівняння вказано деякі

результати, отримані іншими дослідниками. Скорочення CNN у таблиці розшифровується як Convolutional Neural Network – згорткова нейронна мережа, kNN – класифікатор “k найближчих сусідів”, MLP – multilayer perceptron.

Таблиця 1.

Метод	Відсоток помилок
Linear classifier [2]	12
kNN Euclidean [2]	5.0
LeNet4 (CNN) [2]	1.1
LeNet5 (CNN) [2]	0.8
My CNN w/RBF [5]	0.57
PCNC classifier [7]	0.44
SVM w/RBF kernels [8]	0.42
Simard’s Simple CNN [4]	0.4
LeNet6 (CNN) [3]	0.38
My Simple CNN (ця робота)	0.38
Classifier combination [9]	0.35
Huge MLP [10]	0.35

Рис.2. 38 зразків з тестової вибірки, які були неправильно розпізнані

Необхідно зазначити, що найкращі результати (останні два рядки з таблиці 1) отримано немалою ціною. У роботі [9] показано, що комбінація чотирьох принципово відмінних методів розпізнавання рукописних цифр (при цьому один з них – це згорткова нейронна мережа, описана в роботі [4]) може дати результат, кращий, ніж дає кожен з методів сам по собі. У роботі [10] задача розпізнавання рукописних цифр розв’язана методом “грубої сили”. Автори склали програму для 6-шарової мережі типу back propagation з кількістю нейронів у шарах 2500, 2000, 1500, 1000, 500 та 10. Кількість параметрів для навчання цієї мережі становить біля 12 мільйонів. Программа виконувалася на комп’ютері з процесором Intel Core2 Quad 9450 2.66GHz та графічною картою nVidia GTX 280, при цьому інтенсивно використовувалися обчислювальні ресурси графічної карти.

5. ВИСНОВКИ

В даній роботі запропоновано варіант архітектури згорткової нейронної мережі для розпізнавання зображень рукописних символів з лінійним класифікатором. Мережа нараховує чотири шари. Перший та третій шари мережі здійснюють виділення ознак. Другий шар мережі виконує зважене локальне усереднення з зменшенням просторової роздільної здатності. Останній шар мережі є лінійним класифікатором. Навчання мережі здійснюється за методом зворотного розповсюдження помилки, адаптованим до архітектури згорткової нейронної мережі.

Для експериментальної перевірки розпізнавання рукописних цифр використано базу даних MNIST. Отриманий результат близький до найкращого відомого на цей час для бази даних MNIST. Це означає, що проста згорткова нейронна мережа з лінійним класифікатором на виході спроможна ефективно розв’язувати задачу розпізнавання рукописних цифр.

Слід зауважити, що запропонована нейронна мережа має меншу кількість нейронів та відповідно меншу кількість вагових коефіцієнтів зв’язків нейронів, ніж мережі LeNet5, LeNet6 та “simple CNN”, не кажучи вже про мережу з роботи [10]. Це означає, що в разі практичного використання вона могла б забезпечити вищу швидкодію системи розпізнавання.

ЛІТЕРАТУРА

[1] LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., and Jackel, L.D. “Backpropagation applied to handwritten zip code recognition”, *Neural Computation, vol. 1, No 4, p. 541-551, 1989.*

[2] LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. “Gradient-Based Learning Applied to Document Recognition”, *Proceedings of the IEEE, Vol. 86, No. 11, November 1998, p. 2278-2324.*

[3] Bottou, L., Chapelle O., DeCoste D., and Weston J. Large-Scale Kernel Machines. MIT Press, Cambridge, USA, 2007. - 416 pp.

[4] Simard, P.Y., Steinkraus, D., and Platt, J.C. “Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis”, in *Proceedings of the Seventh International Conference on Document Analysis and Recognition*, August 03 - 06, 2003 Edinburgh, Scotland, p. 958-962.

[5] Лукович В. “Згорткова нейронна мережа для розпізнавання рукописних цифр” *Пр. 8-ї міжнар. конф. УкрОбраз’2006*, Київ, 2006. – С.135-138.

[6] Gosselin, B., "Improved Hand-written Character Recognition thanks to a New Geometric Distortion Method", *Proc. of the 6th Int. Conf. On Image Processing and its Applications*, 1997, Dublin, Ireland, vol. 1, p. 327-331.

[7] Kussul, E., Baidyk, T., Wunsch, D. C., Makeyev, O., and Martin, A. “Permutation Coding Technique for Image Recognition Systems”, *IEEE Transactions on Neural Networks. - 2006. – vol. 17, No 6. - p.1566-1579.*

[8] Liu, C.-L., Nakashima, K., Sako, H., Fujisawa, H., “Handwritten Digit recognition: benchmarking of state-of-the-art techniques”, *Pattern Recognition, 36(10): p. 2271-2285, 2003.*

[9] Keysers, D., “Comparison and Combination of State-of-the-art Techniques for Handwritten Character Recognition: Topping the MNIST Benchmark”, *Technical Report, IUPR Research Group, DFKI and Technical University of Kaiserslautern, May 2006.*

[10] Ciresan, D.C., Gambardella, L.M., Schmidhuber, J. Meier, U., “Deep Big Simple Neural Nets Excel on Handwritten Digit Recognition”, available at: http://arxiv.org/pdf/1003.0358v1.