

Інформаційні технології

УДК 004.93

Скороход Іван Григорович

студент

Дніпровського національного університету імені Олеся Гончара

Skorokhod Ivan

Student of the

Oles Honchar Dnipro National University

ВИКОРИСТАННЯ ЗГОРТКОВИХ НЕЙРОННИХ МЕРЕЖ У РОЗПІЗНАВАННІ ТЕКСТІВ

USE OF CONVOLVED NEURAL NETWORKS IN TEXT RECOGNITION

Анотація. В статті проводиться дослідження використання згорткових нейронних мереж при розпізнаванні образів, тексту, звуків, тощо. Приводяться математичні моделі, які використовуються при побудові нейронної мережі та наводяться приклади використання таких моделей у реальному житті. У якості заключення проводиться аналіз можливостей використання нейронних мереж при створенні програмних застосунків.

Ключові слова: нейронні мережі, розпізнавання тексту, глибоке навчання, математичні моделі.

Summary. The article investigates the use of convolutional neural networks in the recognition of images, text, sounds, etc. Mathematical models used in the construction of a neural network are given and examples of the use of such models in real life are given. In conclusion, an analysis of the possibilities of using neural networks in the creation of software applications.

Key words: neural networks, text recognition, deep learning, mathematical models.

Вступ. Сучасний світ дуже сильно переплетений з інформаційними технологіями та інноваціями, величезна кількість компаній та винахідників працюють над створенням чогось нового, що допоможе покращити життя людей або полегшить виробництво для цього використовуються в першу чергу математичні науки. Так на сучасному етапі розвитку людство прийшло до таких технологій як – нейронні мережі. Першопроходцями у цьому напрямку були американські вчені У. Маккалок та У. Путтс.

Нейронні мережі з кожним днем захоплюють все більше і більше ринку, на даний час нейронні мережі використовуються майже у всіх сферах життя, починаючи від авіоніки та сектору безпеки та закінчуючи дитячими іграшками.

Так наприклад у секторі безпеки нейронні мережі отримали неймовірний розвиток. В першу чергу в Китаї та США. В часи епідемії COVID-19 нейронні мережі дуже сильно використовуються у системах відео спостереження для встановлення хворих осіб, а також для контролю осіб які проходять самоізоляцію.

Передові компанії світу такі як, Google, Facebook, Apple, Amazon, Microsoft борються на ринку за найкращих фахівців у сфері машинного навчання. Всі вони бачать неймовірні можливості у використанні нейронних мереж.

Також нейронні мережі та математичні дослідження побудовані на використанні нейронних мереж у неймовірних кількостях використовуються у наукових дослідженнях. Біологічні дослідження, дослідження генному, економічні прогнозування та дослідження, розробка, проектування та створення космічних кораблів.

Постановка проблеми. Необхідність розпізнавання об'єктів у сучасному житті знаходиться на дуже високому рівні, так розпізнавання обличч в правоохоронній сфері, створення безпілотних систем дуже актуальні

на теперішній час. Всі ці системи потребують використання технологій розпізнавання об'єктів, для цього використовуються нейронні мережі.

Неймовірно швидкі темпи зростання кількості та обсягів даних, які людство створює у останній часи у деяких випадках ускладнює пошук необхідної інформації через те, що велика кількість текстів просто сканується у pdf або у формати зображень, тому ця інформація не індексується пошуковими системами. Отрим того при наборі текстів у цифровій формі все ще використовується робота людини і це надає простір для покращень. Щоб за допомогою комп'ютерного розпізнавання одразу переводити текст в цифрову форму.

Найчастіше для розпізнавання зображень та тексту використовуються згорткові нейронні мережі.

Аналіз останніх досліджень і публікацій. Стаття Python + OpenCV + Keras за посиланням <https://habr.com/ru/post/466565/> - хороша стаття з прикладами коду створення нейронної мережі на мові Python, описуються основні бібліотеки, які використовуються при створенні нейронних мереж та машинному навчанні на мові Python. По суті повністю описано створення нейронної мережі розпізнавання тексту з використанням популярних бібліотек.

Стаття Learning Recognition and Segmentation Using the Cresceptron, J. Weng, Dept of Computer Science, Michigan State University, East Lansing, MI 48824 USA хороший теоретичний матеріал на тему візуального розпізнавання зображень за допомогою нейронних мереж. Дуже багато схематичних ілюстрацій, які показують як саме працює нейронна мережа при розпізнаванні образів. Проводяться математичні розрахунки роботи нейронних мереж, надається широка інформація щодо обчислення похибки обчислення нейронних мереж.

Стаття Learning Recognition and Segmentation of 3-D Objects from 2-D Images, J. Weng, Dept of Computer Science, Michigan State University, East Lansing, MI 48824 USA стаття від автора попередньої статті, більш розширений варіант, з більш детальнішим описом процесів розпізнання 3-Д та 2-Д об'єктів. Надана величезна кількість прикладів розпізнавання зображень людського обличчя.

Книга Practical Machine Learning with Python, Dipanjan Sarkar, Bangalore, Karnataka, India. Книга у якій описано повний процес створення нейронних мереж різного ступеня складності та різного застосування. Описано повний теоретичний процес створення та роботи нейронних мереж. Описується екосистема машинного навчання на мові Python. Наявні реальні приклади нейронних мереж починаючи від аналізу музики та завершуючи системи роботи на фондових ринках та біржах.

Мета дослідження. За результатами дослідження потрібно сформулювати теоретичну та математичну базу для побудови нейронної мережі розпізнавання тексту з фото. Провести дослідження роботи згорткових нейронних мереж. Встановити можливості використання згорткових нейронних мереж при розпізнаванні фізичних об'єктів, звуків тощо.

Використання згорткових нейронних мереж у розпізнаванні тексту. Згорткова нейронна мережа (ЗНМ) - це один з видів нейронних мереж для обробки даних з сітчастою топологією. Прикладами можуть бути тимчасові ряди, які можна розглядати як одномірну сітку прикладів, що вибираються через регулярні проміжки часу, а також зображення, що розглядаються як двовимірна сітка пікселів. Згорткові мережі неймовірно сильно розповсюдилися та використовуються у багатьох сферах наукових досліджень та у реальному житті сучасної людини. Своєю назвою вони зобов'язані використанню математичної операції згортки. Згортка - це особливий вид

лінійної операції. Згорткові мережі – це просто нейронні мережі, в яких замість загальної операції множення на матрицю принаймні в одному шарі використовується згортка.

У загальному вигляді згортка – це операція над двома функціями речовинного аргументу. Щоб обґрунтувати визначення згортки, почнемо з прикладів можливих функцій.

Припустимо, що ми стежимо за положенням літака за допомогою лазерного датчика. Наш датчик видає єдине значення $x(t)$, положення літака у просторі в момент часу t . Змінні x і t приймають речові значення, тобто показання датчика будь-які два моменти часу можуть відрізнитися. Тепер припустимо, що у роботі датчика присутні перешкоди. Щоб отримати менш зашумлену оцінку положення літака, необхідно усереднити кілька результатів вимірів. Зрозуміло, нещодавні виміри важливіші, тому ми хочемо обчислювати виважене середнє, надаючи недавнім вимірам більшої ваги. Для цього можна скористатися ваговою функцією $\omega(a)$ де a - давність вимірювань. Застосувавши таку операцію усереднення в кожний момент часу, ми отримаємо нову функцію, яка дає згладжену оцінку положення літака:

$$s(t) = \int x(a)\omega(t-a)da$$

Ця операція називається згорткою і зазвичай позначається зірочкою:

$$s(t) = (x * \omega)(t)$$

У нашому прикладі ω має бути функцією щільності ймовірності, інакше усереднення не вийде. Крім того, ω повинна дорівнювати 0 для всіх негативних значень аргументу, інакше вона буде здатна заглядати в майбутнє, що навряд чи у межах наших можливостей. Але ці обмеження характерні лише для цього прикладу. У загальному випадку згортку можна визначити для будь-яких

функцій, для яких визначено наведений вище інтеграл, і використовувати не тільки для отримання виваженого середнього.

У термінології згорткових мереж перший аргумент (у прикладі функція x) називається входом, а другий (функція ω) – ядром. Вихід іноді називають картою ознак.

Лазерних датчиків, здатних видавати результати вимірів у будь-який момент часу, насправді немає. Зазвичай під час роботи з даними в комп'ютері час дискретизовано, і наш датчик видаватиме дані через регулярні інтервали. Мабуть, було б реалістичніше припустити, що лазер виробляє вимірювання раз на секунду.

Індекс моменту часу t може приймати лише цілі значення. Якщо тепер припустити, що x і ω визначено лише для цілих t , то ми отримаємо визначення дискретної згортки:

$$s(t) = (x * \omega)(t) = \sum_{a=-\infty}^{\infty} x(a)\omega(t - a)$$

У програмах машинного навчання входом зазвичай є багатовимірний масив даних, а ядром – багатовимірний масив параметрів, адаптованих алгоритмом навчання. Будемо називати ці масиви тензорами. Оскільки кожен елемент входу та ядра повинен зберігатися окремо у явному вигляді, зазвичай передбачається, що ці функції дорівнюють нулю всюди, крім кінцевої множини точок, для яких ми зберігаємо значення. Насправді це означає, що суму від мінус до плюс нескінченності можна замінити підсумовуванням по кінцевому числу елементів масиву. Нарешті, ми часто використовуємо пакунки відразу по кількох осях. Наприклад, якщо входом є двовимірне зображення I , то ядро має бути двовимірним:

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(m, n)K(i - m, j - n)$$

Операція згортки коммутативна, тому формулу можна записати і так:

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(i - m, j - n)K(m, n)$$

Зазвичай другу формулу простіше реалізувати у бібліотеці машинного навчання, оскільки діапазон допустимих значень m та n менший.

Властивість комутативності згортки має місце, тому що ми відобразили ядро щодо входу, тобто при збільшенні m індекс входу збільшується, а індекс ядра зменшується. Єдина причина такого відображення – забезпечити комутативність.

І хоча комутативність корисна для доказу теорем, у реалізації нейронних мереж вона зазвичай не грає ролі. Натомість у багатьох бібліотеках реалізована споріднена функція – перехресна кореляція – та ж згортка, тільки без відображення ядра:

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(i + m, j + n)K(m, n)$$

У багатьох бібліотеках машинного навчання реалізовано саме перехресну кореляцію, але називається вона згорткою. Далі ми дотримуватимемося цього ж і називатимемо згорткою обидві операції. У контексті машинного навчання алгоритм навчання ставить знайдені значення ядра у правильну позицію, тому якщо алгоритм заснований на згортці з відображенням ядра, він навчить ядро, відбите щодо того, що навчено алгоритмом зі згорткою без відображення. Рідко буває так, щоб згортка використовувалася в машинному навчанні сама по собі; частіше вона комбінується з іншими функціями, і такі комбінації не комутативні не залежно від того, використовується в ядрі-згортці відображення чи ні.

Дискретну згортку можна як множення на матрицю, на елементи якої накладені деякі обмеження. Наприклад, у разі одновимірної дискретної згортки кожен рядок матриці повинен дорівнювати попередній, зсунутій на один елемент. Це твердження називається теорема Тепліца. У двовимірному випадку згортці відповідає двічі блочно-циркулянтна матриця. Крім обмежень

на рівність деяких елементів, згортці зазвичай відповідає сильно розріджена матриця (у якій більшість елементів дорівнює нулю). Пов'язано це з тим, що ядро, як правило, набагато менше вхідного зображення. Будь-який алгоритм нейронної мережі, заснований на множенні матриць і не залежить від особливостей структури цих матриць, повинен працювати і з згорткою без будь-яких модифікацій самої мережі. У типових згорткових нейронних мережах все ж таки використовуються особливості структури, щоб організувати ефективну обробку великих входів, але з теоретичної точки зору це необов'язково.

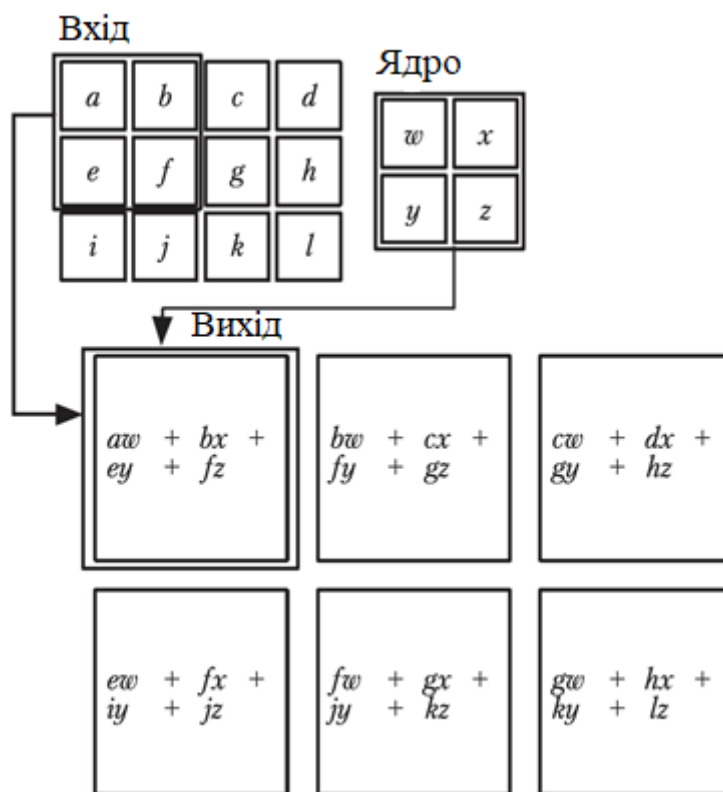


Рис. 1. Приклад двовимірної згортки без відображення ядра. Вихід обмежений лише тими позиціями, котрим ядро цілком укладається у зображення; у деяких контекстах така згортка називається «допустимою». Прямокутник з стрілками, що вказують на нього, показує, як лівий верхній елемент вихідного тензора утворюється шляхом застосування ядра до відповідної лівої верхньої області вхідного тензора

У слоях традиційної нейронної мережі застосовується множення на матрицю параметрів, у якій взаємодія між кожним вхідним та кожним вихідним блоками описується окремим параметром. Це означає, що кожен вихідний блок взаємодіє з кожним вхідним блоком. Навпаки, у згорткових мережах взаємодії зазвичай розріджені (цю властивість називають ще розрідженою зв'язністю, або розрідженими вагами). Досягається це за рахунок того, що ядро менше входу. Наприклад, вхідне зображення може містити тисячі чи мільйони пікселів, але невеликі значущі ознаки, наприклад кордону, можна знайти за допомогою ядра, що охоплює всього десятки чи сотні пікселів. Отже, потрібно зберігати менше параметрів, а це знижує вимоги моделі до обсягу пам'яті та підвищує її статистичну ефективність. Крім того, для обчислення виходу потрібно менше операцій. Все разом зазвичай набагато підвищує ефективність мережі. Якщо є m входів і n виходів, для множення матриць потрібно $m \times n$ параметрів, і складність практично використовуваних алгоритмів становить $O(m \cdot n)$ (в розрахунку на один приклад). Якщо обмежити кількість з'єднань з кожним виходом величиною k , то знадобиться тільки $k \cdot n$ параметрів, і складність становитиме $O(k \cdot n)$. У багатьох практичних додатках можна отримати хорошу якість, коли k на кілька порядків менше m . Графічно розріджена зв'язність проілюстровано на рис. 2 та рис. 3.

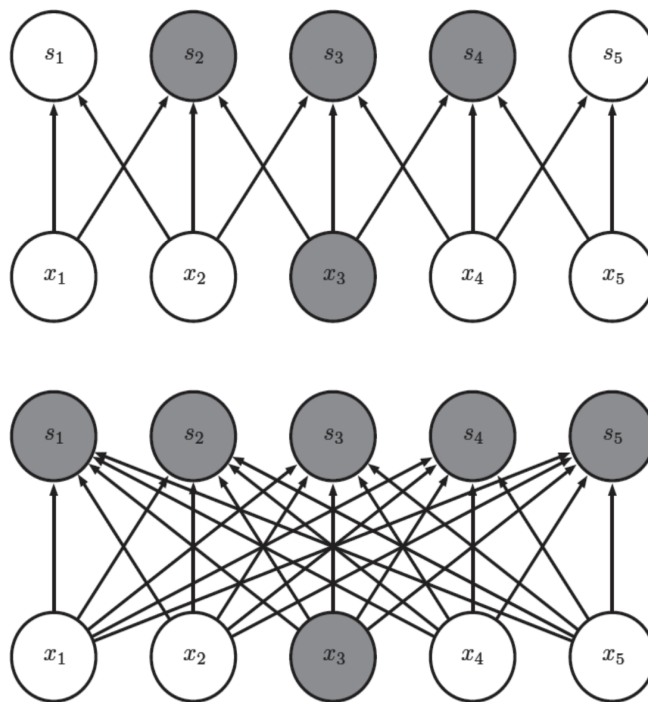


Рис. 2. Розріджена зв'язність при погляді знизу. Виділено один вхідний блок x_3 та вихідні блоки у шарі s , на які цей блок впливає. (Вгорі). Якщо s утворений згорткою з ядром ширини 3, x_3 впливає лише на три виходи. (Внизу) Якщо s утворений множенням на матрицю, то зв'язаність вже не розріджена, тому x_3 впливає на всі виходи

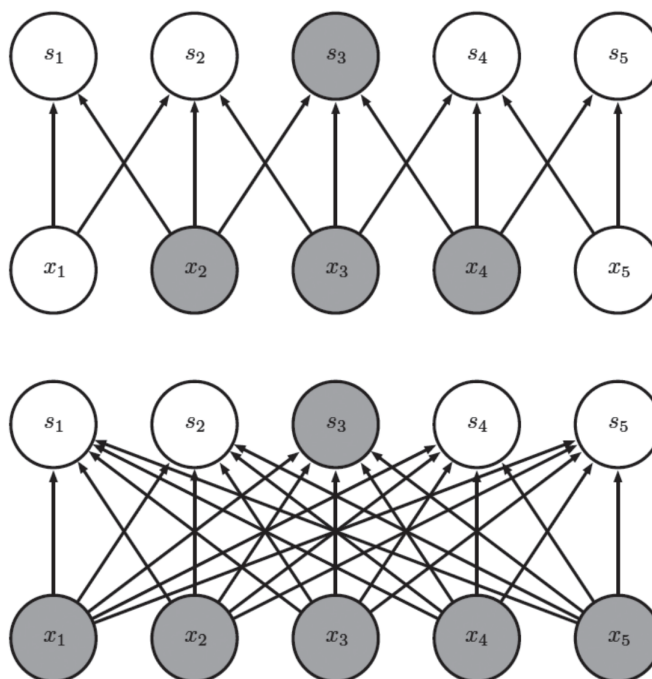


Рис. 3. Розріджена зв'язність при погляді зверху. Виділено один вихідний блок x_3 та вихідні блоки у шарі x , які на нього впливають. Сукупність цих блоків називається рецептивним полем x_3 . (Вгорі) Якщо s утворений згорткою з ядром ширини 3, то на x_3 впливають лише три входи. (Внизу) Якщо s утворений множенням на матрицю, то зв'язність вже не розріджена, тому на x_3 впливають усі входи

У глибокій згортковій мережі блоки нижніх рівнів можуть опосередковано взаємодіяти з більшою частиною мережі, як показано на рис. 4. Це дає мережі можливість ефективно описувати складні взаємодії між багатьма змінними шляхом складання з простих будівельних блоків, кожен із яких описує лише розріджені взаємодії.

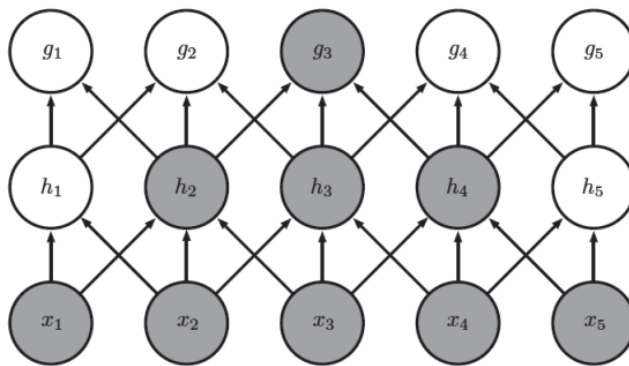


Рис. 4. Рецептивне поле блоків у глибоких шарах згорткової мережі більше рецептивного поля у шарах, близьких до поверхні. Цей ефект посилюється, якщо мережа включає такі архітектурні особливості, як згортка з кроком або пулінг. Це означає, що хоча прямі зв'язки в мережі згортки дійсно дуже розріджені, блоки у глибоких шарах можуть бути опосередковано пов'язані з усім вхідним зображенням або з більшою його частиною

При дослідженні згортки в контексті нейронних мереж ми зазвичай не маємо на увазі стандартну операцію дискретної згортки в тому сенсі, в якому вона розуміється у математичній літературі. Функції, що практично використовуються, трохи відрізняються.

Насамперед під згорткою в нейронних мережах ми зазвичай розуміємо операцію, що складається з багатьох паралельних обчислень згортки. Пов'язано це з тим, що згортка з одним ядром може виділити лише один вид ознак, хоча й у багатьох місцях. Зазвичай хочемо, щоб кожен шар мережі виділяв багато різних ознак у багатьох місцях.

Крім того, вхід зазвичай є не просто сіткою речових значень, а сіткою векторні спостереження. Наприклад, кожен піксель кольорового зображення складається з червоної, зеленої та синьої компонент. У багат шаровій згортковій мережі входом другого шару є вихід першого, який зазвичай складається з результатів багатьох різних згорток у кожній позиції.

Припустимо, що є чотиривимірний тензор K , елемент K_{ijkl} якого визначає силу зв'язку між блоком i – го вихідного каналу та j – го вхідного каналу, коли вхідний та вихідний блоки відстають один від одного на k строк та l стовбців. Припустимо також, що вхід складається з даних спостережень V , де елемент V_{ijk} містить значення i – го каналу вхідного блоку на перетині j – го рядка та k – го стовпця. І нехай вихід Z має такий самий формат, як V . Якщо Z породжений згорткою V зі ковзним ядром K без відображення K , то:

$$Z_{ijk} = \sum_{l,m,n} V_{l,j+m-1,k+n-1} K_{i,l,m,n}$$

де підсумовування проводиться у всіх індексах l, m, n , для яких елементи тензорів під знаком суми існують.

Для зменшення вартості обчислень деякі положення ядра іноді пропускають (ціною зниження точності виділення ознак). Ми можемо розглядати це як знижуючу передискретизацію виходу повної функції згортки. Якщо ми хочемо вибирати тільки кожен s -ий піксель у кожному напрямі виходу, то можемо визначити функцію згортки зниженого розширення:

$$Z_{ijk} = c(K, V, s)_{i,j,k} \sum_{l,m,n} [V_{l,(j-1)xs+m,(k-1)xs+n} K_{i,l,m,n}]$$

Будемо називати s кроком згортки зниженого дозволу.

Дані, що використовуються в мережі згортання, зазвичай складаються з декількох каналів, кожен з яких містить спостереження якоїсь величини у певній точці простору або у певний момент часу. Приклади типів даних різної розмірності з різним числом каналів наведено у табл. 1.

Таблиця 1

Приклади форматів даних у згорткових мережах

	Одноканальні	Багатоканальні
1D	Аудіосигнали: вісь згортки відповідає часу. Ми дискретизуємо час та вимірюємо амплітуду сигналу один раз у кожному часовому інтервалі	Дані анімації «скелета»: анімації тривимірного малювання комп'ютером персонажа генеруються шляхом зміни положення "скелета" з часом. Кожен момент положення скелета описується кутами зчленування кісток у кожному суглобі. Кожен канал даних, що подаються на вхід згорткової мережі, представляє кут відносно однієї осі одного суглоба
2D	Аудіодані, попередньо оброблені за допомогою перетворення Фур'є. Ми можемо перетворити аудіосигнал на двовірний тензор, рядки якого відповідають частотам, а стовпці – моментам часу. Застосування пакунка за часом робить модель еквіваріантною щодо тимчасових зрушень. Застосування згортки по осі частот робить модель еквіваріантною щодо частоти, тобто одна і та ж мелодія в різних октавах породжує на вході мережі одне й те саме уявлення, але з різною висотою	Дані кольорового зображення: один канал містить червоні пікселі, інший – зелені, третій – сині. Ядро згортки зсувається по двох осях зображення, забезпечуючи еквіваріантність щодо паралельного перенесення в обох напрямках
3D	Об'ємні дані: типовим джерелом таких даних є технології медичної інтроскопії, наприклад, комп'ютерної томографії	Дані кольорового відео: одна вісь відповідає часу, інша – висоті кадру, третя – ширині кадру

У сучасних додатках згорткових мереж часто беруть участь мережі, що містять понад мільйон блоків. Для роботи з ними необхідні ефективні реалізації, засоби розпаралелювання обчислень. Але в багатьох випадках роботу можна прискорити, вибравши відповідний алгоритм згортки.

Пошук більш швидких способів обчислити згортку точно або приблизно, не приносячи в жертву вірності моделі, область активних досліджень. Навіть методи, що покращують ефективність одного лише прямого поширення, вже корисні, тому що в комерційних системах зазвичай більше ресурсів витрачається не так на навчання, але в розгортання мережі.

Згорткові мережі відіграли важливу роль в історії глибокого навчання. Це яскравий приклад успішного застосування ідей, висловлених у процесі вивчення людського мозку, до машинного навчання. Заодно вони виявилися одними з перших глибоких моделей гарної якості, задовго до того, як глибокі моделі взагалі були визнані життєздатними. Згорткові мережі були також одними з перших нейронних мереж, що знайшли застосування у важливих комерційних додатках, і досі вони залишаються у перших рядах комерційних застосувань глибокого навчання. Наприклад, у 1990-ті роки дослідницька група з нейронних мереж у компанії AT&T розробила мережу для розпізнавання чеків. До кінця 1990-х років ця система була розгорнута компанією NEC і розпізнавала 10% усіх чеків у США. Пізніше кілька систем розпізнавання символів і рукописних текстів на основі мереж згортання було розгорнуто корпорацією Microsoft.

Згорткові мережі були одними з перших працюючих глибоких мереж, навчених із застосуванням зворотного поширення. Не цілком зрозуміло, чому згорткові мережі досягли успіху там, де мережі загального призначення зі зворотним розповсюдженням зазнали невдачі. Можливо, справа в тому, що згорткові мережі обчислювально ефективніші за повнозв'язні мережі, тому було простіше ставити експерименти і оптимізувати реалізацію та гіперпараметри. Крім того, що більше мережа, то легше її навчити. На сучасному обладнанні великі пов'язані мережі дають непогані результати на багатьох завданнях навіть при використанні наборів даних та функцій

активації, які були доступні та популярні в ті часи, коли вважалося, що від повнозв'язних мереж неможливо досягти хорошої якості. Можливо також, що основні бар'єри на шляху до успіху нейронних мереж були суто психологічної властивості (фахівці-практики не очікували від них гарної якості, а тому не особливо намагалися їх використати). Як би там не було, нам пощастило, що згорткові сітки добре працювали кілька десятиків років тому. У багатьох сенсах вони передали естафету іншим розділам глибокого навчання та проклали шлях прийняттю нейронних мереж загалом.

Згорткові мережі дозволяють спеціалізувати нейронні мережі для роботи з даними, що мають чітко виражену топку сітки, і добре масштабувати такі моделі до завдань дуже великого розміру. Особливо успішним цей підхід виявився у застосуванні до двовимірних зображень.

Висновки. В цій науковій роботі було проведено дослідження у напрямі створення цифрових нейронних мереж, та їх використання при машинному розпізнаванні тексту.

Було опрацьовано та створено математичну модель нейронної мережі, описано формули які використовуються при створенні нейронних мереж для розпізнавання тексту.

Було розглянуто та досліджено архітектуру створення згорткових нейронних мереж. Було проведено теоретичний аналіз роботи згорткових нейронних мереж.

Література

1. URL: <https://habr.com/ru/post/466565/>, стаття з сайту.
2. Стаття Learning Recognition and Segmentation Using the Cresceptron, J. Weng, Dept of Computer Science, Michigan State University, East Lansing, M1 48824 USA.

3. Стаття Learning Recognition and Segmentation of 3-D Objects from 2-D Images, J. Weng, Dept of Computer Science, Michigan State University, East Lansing, M1 48824 USA.
4. Книга Practical Machine Learning with Python, Dipanjan Sarkar, Bangalore, Karnataka, India.
5. Книга Machine Learning with Python, tutorials point.

References

1. URL: <https://habr.com/ru/post/466565/>, article from the site.
2. Article Learning Recognition and Segmentation Using the Cresceptron, J. Weng, Dept. of Computer Science, Michigan State University, East Lansing, M1 48824 USA.
3. Article Learning Recognition and Segmentation of 3-D Objects from 2-D Images, J. Weng, Dept of Computer Science, Michigan State University, East Lansing, M1 48824 USA.
4. Book Practical Machine Learning with Python, Dipanjan Sarkar, Bangalore, Karnataka, India.
5. Book Machine Learning with Python, tutorials point.