

УДК 004.855.5

Нейронні мережі в сучасну епоху

І.В. Гущин, В.М. Куклін, О.В. Мішин

Гущин Іван Валерійович *старший викладач, кафедра штучного інтелекту та програмного забезпечення, Харківський національний університет імені В. Н. Каразіна, майдан Свободи 4, м. Харків-22, Україна; 61022*
e-mail: i.v.gushchin@karazin.ua, <http://orcid.org/0000-0002-1917-716X>

Куклін Володимир Михайлович *доктор фізико-математичних наук, професор; завідувач кафедри штучного інтелекту та програмного забезпечення, Харківський національний університет імені В. Н. Каразіна, майдан Свободи, 4, Харків-22, Україна, 61022*
e-mail: v.m.kuklin@karazin.ua; kuklinvm1@gmail.com; <https://orcid.org/0000-0002-0310-1582>

Мішин Олександр Вікторович *старший викладач, кафедра штучного інтелекту та програмного забезпечення, Харківський національний університет імені В. Н. Каразіна, майдан Свободи 4, м. Харків-22, 61022, Україна;*
e-mail: o.v.mishyn@karazin.ua, <http://orcid.org/0000-0001-7478-757X>

Показана ідея застосовувати уявлення про роботу людського мозку як математичні моделі, які можна використовувати в безлічі прикладних задачах. Показано, що математичні та фізіологічні моделі по суті мають досить мало спільного, але основну ідею - застосовувати нейрони, як деякі самостійні обчислювальні вузли і комбінувати їх в шари - залишили і розвинули. Наведено історичні підґрунтя розвитку науки нейронних мереж, згадані основні дослідники, чії роботи впливали на вектор і темп розвитку технологій найсильніше. Показано, що нейронні мережі мали різну підтримку з боку інвесторів за час свого розвитку і пік масового інтересу до них був залежний від виникнення необхідних обчислювальних потужностей або проривний архітектури мережі. Такими були свого часу перцептрони, мережі з наявністю зворотного зв'язку, мережі, які застосовують операцію згортки для аналізу і класифікації зображень. Показано, що так зване глибоке навчання розвинулося на базі методів оптимізації ваг градієнтним спуском. Проведено огляд відомих рішень мереж навчання без учителя, зі зворотним зв'язком і мовні моделі. Генеративні моделі виглядають, як найбільш перспективний напрямок у розвитку наукової думки і створенні інтерпретованих рішень на базі нейронних мереж.

Ключові слова: нейронні мережі, перцептрон, навчання з учителем, навчання без учителя, глибоке навчання, зворотне поширення помилки, згортка матриць, поширене навчання, градієнтний спуск, функція активації.

Neural Networks in the Modern Age

I. Gushchin, V. Kuklin, A. Mishyn

Gushchin Ivan *senior teacher, Department of Artificial Intelligence and Software V. N. Karasin Kharkiv National University, Svobody Sq 4, 61022, Kharkiv, Ukraine,*

Kuklin Volodymyr *PhD, Doctor of Science, Professor, Head of Department of Artificial Intelligence and Software V. N. Karasin Kharkiv National University, Svobody Sq 4, 61022, Kharkiv, Ukraine;*

Mishyn Alex *senior teacher, Department of Artificial Intelligence and Software V. N. Karasin Kharkiv National University, Svobody Sq 4, 61022, Kharkiv, Ukraine,*

The idea to apply representations about the work of the human brain as mathematical models, which can be used in a variety of applied tasks is shown. It is shown that mathematical and physiological models have quite little in common, but the basic idea - to apply neurons as some independent computational nodes and combine them into layers - has been developed to the current state of technology. The historical background of the development of neural network science is given and the main researchers, whose works have influenced the vector and the pace of technology development the most, are mentioned. It is shown that

neural networks had different support from investors during their development and the peak of mass interest in them depended on the emergence of necessary computing power or a breakthrough network architecture. Such networks were in its time perceptrons, networks with feedback, networks applying convolution operation for image analysis and classification. It is shown that the so-called deep learning has developed based on weight optimization methods by gradient descent. A review of known solutions of supervised learning, feedback, and language-based learning networks is conducted. Generative models look like the most promising direction in the development of scientific thought and the creation of interpretive solutions based on neural networks.

It is shown that in learning with a teacher, which is typical for deep neural networks, the optimization for it is performed by regularization procedures, which help to avoid diversity and provide error minimization during error back propagation.

Keywords: *neural networks, perceptron, learning with a teacher, learning without a teacher, deep learning, error back propagation, matrix convolution, layer-by-layer learning, gradient descent, activation function.*

Нейронные сети в современную эпоху

И.В.Гущин, В.М.Куклин, А.В.Мишин

Гущин Иван Валерьевич

старший преподаватель,, кафедра искусственного интеллекта и программного обеспечения, Харьковский национальный университет имени В. Н. Каразина, площадь Свободы, 4, Харьков-22, Украина, 61022

Куклин Владимир Михайлович

доктор физико-математических наук, профессор; заведующий кафедры искусственного интеллекта и программного обеспечения, Харьковский национальный университет имени В. Н. Каразина, площадь Свободы, 4, Харьков-22, Украина, 61022

Мишин Александр Викторович

старший преподаватель,, кафедра искусственного интеллекта и программного обеспечения, Харьковский национальный университет имени В. Н. Каразина, площадь Свободы, 4, Харьков-22, Украина, 61022

Показана идея применять представления о работе человеческого мозга как математические модели, которые можно использовать во множестве прикладных задачах. Показано, что математические и физиологические модели по сути имеют достаточно мало общего, но основную идею – применять нейроны, как некоторые самостоятельные вычислительные узлы и комбинировать их в слои – оставили и развили. Приведены исторические подоплеку развития науки нейронных сетей, упомянуты основные исследователи, чьи работы влияли на вектор и темп развития технологий сильнее всего. Показано, что нейронные сети имели различную поддержку со стороны инвесторов за время своего развития и пик массового интереса к ним был зависим от возникновения необходимых вычислительных мощностей или прорывной архитектуры сети. Такими были в свое время перцептроны, сети с наличием обратной связи, сети, применяющие операцию свертки для анализа и классификации изображений. Показано, что так называемое глубокое обучение развилось на базе методов оптимизации весов градиентным спуском. Проведен обзор известных решений сетей обучения без учителя, с обратной связью и языковые модели. Генеративные модели выглядят, как наиболее перспективное направление в развитии научной мысли и создании интерпретируемых решений на базе нейронных сетей.

Ключевые слова: *нейронные сети, перцептрон, обучение с учителем, обучение без учителя, глубокое обучение, обратное распространение ошибки, свертка матриц, послойное обучение, градиентный спуск, функция активации.*

1 Вступ

Зв'язок логіки і нервової діяльності підказала Мак-Каллоку У.С. і Піттсу В. в 1943 р ідею нейронної мережі [1]. Поява книги «Кібернетика, або управління і зв'язок в тварині і машині» в 1948 р. Н. Вінера, ознаменувало появу нової галузі знання - кібернетики. Здавалося, ці два напрямки повинні об'єднатися, що і сталося спочатку, але подальший розвиток теорії нейронних мереж У. Мак Каллоком призвело до відмови від об'єднання цієї теорії з кібернетикою.

Д. Хебб вже в 1949 р представляє алгоритми навчання нейронної мережі, але лише в 1958 році з'являється перцептрон Ф. Розенблатта. В 1960 році Б. Уїдроу надали Адалін - адаптивний суматор (використовується в системах обробки сигналів) на керованих резисторах-мімісторах.

Математики відзначили, що вид функції активації, яка передбачається бути обмеженою і нелінійною, якісно не впливає на характеристики мережі¹ для отримання потрібних рішень. В значно більшій мірі важливі зв'язки між нейронами. На відміну від природної нейронної мережі, штучна мережа, взагалі кажучи, не передбачає змін числа і стану ваг окремих нейронів,

¹ Спираючись на роботу Стоуна, Хехт-Нільсон довів, що функції загального вигляду представимо в двошаровій мережі з прямими повними зв'язками [2]

форми передавальної функції і умов на її вході² вже в процесі вирішення завдань. Один нейрон дає можливість обчислити значення однієї нелінійної функції (передавальної), безліч нейронів може обчислювати значення суперпозиції таких функцій. Послідовно спрощуючи вид уявлення довільної функції як твори і суперпозиції стандартних функцій спочатку А. Н. Колмогоров [3], потім Г. Лоренц [4], а потім і Д. Шпрехер [5] прийшли до висновку, що будь-яка функція може бути представлена як кінцева сума деяких функцій, аргументом яких є кінцеві суми стандартних функцій. Іншими словами, шари мережі при певній її архітектурі формують суму своїх відгуків (кожен відгук у вигляді цієї деякої функції), аргументами якої є функції активації окремих нейронів. Математики підказали програмістам, що таким чином можна отримати в результаті будь-яку функцію, тобто мережа дозволяє розраховувати, що вона знайде рішення в будь-якому випадку³.

Згодом оптимізм згас. Невдачі машинного перекладу (в основному з російської на англійську) змусили розчарованих керівників США в період холодної війни згорнути фінансування цієї програми. Удачі логічних методів (символьні підходи - *symbolic reasoning*) штучного інтелекту відсунули на узбіччя підтримки чиновників прихильників коннективізму (винахідливого створення блоків з нейронів, як в мережі У. Питтса і У. Мак-Каллока) і успіх перцептронів. Сумніви А. П. Петрова і М. М. Бонгарда [7,8] в можливостях перцептрона в 1963р., а також робота однокурсника Ф. Роземблатта М. Мінського про неможливість вирішувати завдання з інваріантними уявленнями (див. також проблему єдиного рішення) в 1969 р, привели до першої зими в розвитку нейронних мереж. Хоча в цей час інтерес до цієї теми не пропав (див., наприклад, [9,10]). Т. Кохонен і Дж. Анлдерсон в 1972 р запропонували використовувати нейронні мережі як системи пам'яті. Б. В. Хакимов 1973 р використовували модель з синапсами для вирішення завдань, П. Дж. Веброс, а також А. І. Галушкін 1974 р обговорювали алгоритм зворотного поширення помилки, але ці роботи залишилися без уваги. Навіть ідея мережі, що самоорганізується - когнітрону і потім неокогнітрона Фукусіми (Fukushima, 1975, 1980) для розпізнавання образів не викликала інтересу, хоча через два десятиліття лягла в основу сучасної сверточної мережі. Чималу частку розчарування викликав негативний звіт Дж. Лайтхілла про можливість нейронних мереж. У цей період мільярдна галузь штучного інтелекту залишилась без фінансування через розчарування чиновників від науки.

Закінчення першої зими ознаменувало появу в 1982 р мережі з зворотними зв'язками Дж. Хопфілда і мережі Кохонена, причому остання навчалася без учителя і забезпечувала кластеризацію рішень самостійно. Потім в 1986 р ціла когорта дослідників (західних - Д. І. Румельхарт, Дж. Е. Хінтон, Р. Дж. Вільямс практично одночасно з вченими красноярської групи на чолі з С. І. Барцевим) повернулися до методу зворотного поширення помилки. Мережі спочатку були невеликими (кількість шарів, кількість нейронів, кількість синапсів-зв'язків), і тому не припинялися спроби створення їх формального - аналітичного опису, що пізніше призвело-таки до створення безлічі теорій глибокого навчання. У цей період стали домінуючими в усвідомленні проблем ролі нейрофізіологів і математиків. Перші, спираючись на знання систем головного мозку, пропонували все нові варіанти архітектури та способи конструювання рішень, а математики намагалися довести раціональність деяких, що сподобалися їм мереж і методів.

Підхід до підтримки досліджень завжди був традиційний для чиновників, особливо військових. Як відомо, агентство передових оборонних досліджень «ARPA» (тепер «DARPA») дало можливість М. Мінських, Г. А. Саймону, А. Ньюеллу і іншим авторитетам витратити будь-які засоби для прогресу в тих областях, які вченим здавалися перспективними. Але вже в 1969 р (поправка Менсфілда) від вчених і технологів зажадали вже цільових розробок. Звіти Дж. Лайтхілла та інших залучених військовими експертів привели з 1974 року в розпал економічної кризи до очікуваного згортання фінансування програм штучного інтелекту (ШІ). Після десятиліття черговий зими досліджень, в 1983 році DARPA запустили Ініціативу стратегічних обчислень (Strategic Computing Initiative). Програмою досліджень керувало Управління технологій обробки інформації (ІРТО). Через чотири роки, розчаровані результатами чиновники все згорнули, крім систем управління боєм DART, де, до речі, штучний інтелект проявив себе прекрасно під час першої війни в Перській затоці 1991 р. Відомим став скандальний факт

² У разі настройки і навчання мережі такі зміни можливі

³ Детальніше про математичній базі нейронних мереж см. [6].

висловлювання колеги вищезазначених вчених Х. Моравека, який звернув увагу громадськості на схильності вчених до нереалістичних прогнозів, а чиновників до завищеної вимогливості до результативності наукового пошуку. Агентство в цей же час також скасувало фінансування дослідникам з університету Карнегі-Меллона, які працювали над програмою розпізнавання мови (на основі прихованих марківських моделей), але тільки при проголошенні слів в певному порядку. Але, як водиться, розробки цих вчених допомогли розвинути ці методи лише через роки, до початку нового століття.

2 Нейронні мережі в XXI столітті

Ще раз про природний інтелект. Для пояснення змін в динаміці розвитку нейронних мереж корисно повернутися до природного інтелекту. Мозок - це нейронна мережа, яку можна з певними умовностями представляти, як багатопшарову структуру з нейронів (хоча в дійсності уявлення у вигляді шарів навряд чи відповідає реальності). На що слід звернути увагу, так це на те, що ці шари (або локальні масиви нейронів, що одне й те саме) окремо, вибірково не піддаються навчанню. Навчання проводиться відразу по всьому масиву нейронів або, принаймні, в значній його частині. Хоча відомо, що різні ділянки мозку відповідають за вироблення різних типів рішень⁴ бо окремі органи чуття підключено до різних досить великих ділянок мозку і зокрема його кори і процес навчання локалізовано саме там. Точно також деякі ділянки мозку пов'язані з руховим апаратом, вегетативною нервовою системою і т.п. Але якщо виключити це, то навряд чи в корі мозку наявна чітко організована пошарова переробка інформації, швидше за все формування рішень відбувається одночасно у великих масивах нейронів. А ось той факт, що навчальна вибірка у людини надзвичайно велика, є визначальним. Бо навчається людина довго.

Поява великих нейронних мереж з мільярдами нейронів, безпрецедентне збільшення швидкодії і неймовірне розширення обсягів пам'яті дозволило повернутися до напівзабутих мереж прямого поширення, використовувати систему навчання з учителем, від якої всі відвернулися, на базі зворотного поширення помилки. Тобто діяти досить прямолінійно. Бо з такими новими можливостями інтелектуальних обчислювальних систем можна було забути про багато колишніх хитрощів, які були вимушеними в умовах слабкої технічної бази. Ці нові технічні та обчислювальні можливості дозволили збільшити навчальні вибірки до багатьох тисяч прикладів. Багато експертів відзначають, що якщо таких прикладів буде кілька мільйонів, то здібності до навчання великих нейронних мереж перевершать можливості людини. Кількість нейронів в мережах подвоювалось кожні два з половиною роки, відбувалось також прискорене зростання числа зв'язків між ними-синапсів. Скоро можливості нейронних мереж і природного розуму зрівняються.

Поява теорій глибокого навчання. Успіхи в області створення і навчання великих нейронних систем породили нові мотиви в розвитку наукових напрямків - теорії так званого глибокого навчання для систем в основному прямого поширення (стали застосовувати мережі з згортковими шарами, де використовується операція матричної згортки для формування ознак зображення у задачах класифікації⁵).

Так як людина освоює світ поступово від простих речей до більш складних, не вимагаючи спочатку формального опису усвідомленого, то цей принцип вирішили застосувати і для навчання систем штучного інтелекту. Важливо, що, усвідомивши прості речі, людина використовує це знання для створення уявлень про речі більш складні. Формуючи при цьому мимоволі ієрархічну систему усвідомлення дійсності. Цей ієрархічний, поетапний підхід навчання стали називати глибоким навчанням. Слід зауважити, що далеко не завжди існує можливість отримати результат, що інтерпретується, з даних, взятих в проміжних шарах мереж глибокого навчання, але іноді цього не потрібно. Практика створення штучного інтелекту зіткнулася з проблемою внесення знань в прилади. Логічні системи отримання рішень зажадали

⁴ Також відомо, що при необхідності, наприклад, при травмах, мозок цілком здатний передати функції ураженої ділянки іншим, уцілілим при травмах областям кори. У нормальних умовах вибір функцій окремих локальних центрів прийняття рішень, мабуть, більше пов'язаний з шляхами надання інформації від органів чуття. І визначається зв'язками з виконавчими структурами, наприклад, руховим апаратом організму.

⁵ Більшість прикладних задач класифікацій можна представити як класифікацію зображень, а цю задачу краще розв'язує архітектура на базі згорткових шарів

величезних обсягів роботи по формуванню баз знань і даних, що здалося людям, охочим створити інтелект порівняний з природним, тупиковим шляхом (див., наприклад, обговорення цього питання [11]). Хоча для багатьох обчислюваних задач це шлях знайшов своє застосування. Труднощі освоєння машиною знань про світ (рішення так званих необчислюваних завдань) було ускладнено впливом безлічі зовнішніх факторів, впливів на характер одержуваної інформації. Це явище мінливості однотипної інформації назвали її варіативністю. Перші нейронні мережі та обчислювальне середовище їхнього розміщення, володіючи порівняно нижчими технічними можливостями в порівнянні з нинішніми, не дозволили творцям цих приладів підступитися до подібних завдань.

Після першої хвилі нейронних систем на основі ідей нейрофізіологів⁶, коли навчали мало не кожен нейрон простих мережевих конструкцій типу одношарового перцептрона і після першої зими розчарування сформувалася друга (коннекціоністська) хвиля активності з технологіями навчання мереж⁷. Але завищені очікування інвесторів середини 90 років минулого століття до кінця першого десятиліття наступного століття знову заморозили інтерес до нейронних мереж⁸. Оживив інтерес до них канадський інститут перспективних досліджень (Canadian Institute for Advanced Research - CIFAR) профінансувавши програму нейронних обчислень і адаптивного сприйняття (Neural Computation and Adaptive Perception - NCAP). Алгоритми навчання вимагали великих обчислювальних потужностей, що вже до середини першого десятиліття було досягнуто. Дж. Хінтон з співавторами продемонстрував як в теорії можна було б ефективно навчати мережу за допомогою стратегії пошарового переднавчання). У 2007 р Дж. Хінтон використовував обмежену машину Больцмана (RBM - Restricted Boltzmann Machine) для навчання⁹.

Таким чином Дж. Хінтон з колегами спровокував сплеск інтересу до інтелектуальних мереж і почався третій етап розвитку цього науково-технологічного напрямку - глибоке навчання. Спираючись традиційно на досвід вивчення світу людиною вони вважали, що можна створити багатошарову мережу, кожен наступний шар якої, спираючись на здобуті знання попереднім шаром, виділяв би нові риси усвідомлюваного процесу [13]¹⁰. Аналітична теорія і програмні форми такого навчання засновані були на виставі, що мережі, що складаються з безлічі шарів можуть навчатися послідовно. Кожен шар спирається на дані попереднього і розширює уявлення про об'єкти опису. Тобто один шар описує просту схему - каркас об'єкта, наступний шар її доповнює і ускладнює і т.д. Це подання було реалізовано в формалізованих моделях опису і мало служити моделлю реальних систем отримання рішень в багатошарових і багатоеlementних системах. Такі уявлення також були інтуїтивно засновані на відомому факті кращого засвоєння нового матеріалу людиною на основі його попереднього знання (досвіду). Хоча це попереднє знання було освоєно в інших умовах і в інший час, але на це вже не звертали уваги. Іншими словами виникнення глибокого навчання, як представлення структури рішень у формі ієрархії

⁶ Вже на цьому етапі нейрофізіологи вичерпали запас ідей, який допомагав розвивати мережеві варіанти штучного інтелекту. Вважали, що це було пов'язано з недостатнім знанням роботи мозку. Хоча їхні експерименти дозволили усвідомити, що рішення різних завдань нейронами є однотипним в межах всього мозку (Von Melchner 2000).

⁷ Цей підхід спирався на уявлення, що процес роздумів може бути модульовано мережею з простих елементів зі зв'язками (нейрони і синапси; слова і семантичні ознаки а також їх зв'язки). Взагалі кажучи, коннекціонізм - це породження когнітивістики - вивчення процесу пізнання, що об'єднує різні дисципліни.

⁸ На тлі успіхів ядерних і графічних методів опису.

⁹ Джеффри Хінтону репортери раніше приписували наступне зауваження: «На більшості конференцій говорять про введення невеликих змін замість того, щоб гарненько подумати і запитати себе: «Чому те, що ми робимо зараз, не виходить? З чим це пов'язано? Давайте зосередимося на цьому ». Це навчання стали пізніше називати глибоким. Для цього використовували мережі з багатьма шарами (зокрема, в сучасних гаджетах і пошукових системах для розпізнавання осіб) до десятка. До речі, в гаджетах також використовували паралельні обчислення за технологією CUDA на графічних картах (GPU - graphics processing unit) [12]. Для стільникового зв'язку пізніше були створені спеціальні процесори - NPU (neural processing unit) і прискорювачі штучного інтелекту (AI accelerator).

¹⁰ Процес такого формального послідовного усвідомлення об'єкта представлений в [13]. За цими уявленнями, перший (видимий) шар формує на основі яскравих пікселів зображення його межі. Другий (прихований) шар, спираючись на ці дані, визначає кути і контури. Третій може на цій основі визначати мікроструктуру зображення і т.д.

вкладених концепцій, як часто вважають зросло із спроб використання аналогій з області нейрофізіології, потім конвекціоністського підходу, а потім стало наслідком зусиль Дж. Хінтона з колегами, які з 2006 року стали навчати мережу за допомогою стратегії пошарового навчання.

Далі уявлення про глибоке навчання розширилося на будь-які багаторівневі системи, кожен рівень яких навчається з використанням отриманих попередніми рівнями знань. Хоча досягнення перших хвиль розвитку і результати освоєння нейронних мереж, вплинули на формальні алгоритми цього глибокого навчання¹¹. Однак зараз теорія глибокого навчання використовує і інші досягнення науки: лінійну алгебру, теорію ймовірностей, теорію інформації, чисельну оптимізацію і багато іншого. Теорії глибокого навчання використовують також нові ідеї - так зване розподілене представлення - групи нейронів навчаються своїми функціями, використовуючи інші дані, що містять потрібні для цієї категорії якості¹². Раніше перспективні методи навчання без вчителя стали замінювати на навчання з учителем. Повернулися до напівзабутого зворотного поширення помилки для навчання, розвинули теорію моделювання нейронними мережами послідовностей¹³, з'явилися мережі з довгої короткостроковою пам'яттю (long short-term memory - LSTM).

Можливо декларовані уточнення шуканого уявлення і формування рішень в послідовних шарах великої мережі корисні для пояснення широкому загалу принципів глибокого навчання. Однак не важко побачити слабкість цього підходу в разі нейронних систем з елементами згорткових, рекурентних і рекурсивних (тим більше, що за загальним визнанням саме такі мережі показують кращі результати і навчання, і застосування) блоків, і тим більше в разі природного інтелекту, де висновок формується вже в загальному масиві нейронів і частково відображений в вихідному (або призначеному, або обраному вихідним) шарі. На базі рекурентних моделей розвинулися сучасні архітектури, які цілком успішно справляються із завданням генерації тексту і зображення (наприклад, GAN моделі, Bert або GPT). Однак в так званих залишкових мережах (ResNet) посилають один і той же сигнал на різні шари, комбінуючи потім пропущений через різні шари результат, що вже не зовсім підходить під вихідну фабулу пошарового навчання. Тобто такі мережі не зовсім підходять під класичну термінологію мереж прямого поширення, де один і той же сигнал послідовно проходить по верствам до виходу. У свою чергу в мережах на кшталт GAN, комбінується LSTM і згорткові шари, тому що генеративні моделі повинні генерувати нове на базі якогось зразка. Особливо в відео контенті шари зі зворотним зв'язком і з пам'яттю справляються краще. Очевидно, що грамотна комбінація кращих інструментів може давати хороші результати. Мабуть тому апологети формальних теорій глибокого навчання намагаються відійти від проблем опису великих нейронних мереж і тим більше природного інтелекту, захоплюючись розвитком нових підходів, навіяних і стимульованих успіхами навчання великих нейронних систем.

Однак розробники нейронних мереж все більше уваги стали приділяти проблемам попередньої обробки і введення великих обсягів інформації на входи нейронних мереж. Найбільш важливою проблемою виявилася обробка мови і текстів природної мови, для чого потрібно застосовувати великі розмічені набори даних при навчанні.

3 Обробка текстів природної мови

Відволікаючись від з'ясування природи формування рішень всередині мережі, звернемо увагу на важливі проблеми введення інформації та її перетворення в форму, зручну для сприйняття мережею [14]. Ці проблеми дещо послабили (можливо незаслужено) інтерес практиків до застосування результатів теорій глибокого навчання. Незважаючи на виявлені здібності мереж до узагальнення, розробники і технологи все більше використовували спеціально підготовлені і розмічені набори, бази даних при навчанні нейронних мереж. Мережа, як правило, добре сприймає речові вектори даних, які зазвичай подаються на зовнішній вхідний шар і

¹¹ Так модель ADALINE могла передбачати числа виходячи з вхідних даних. Модифікований її варіант використовується в підходах глибокого навчання.

¹² Навчання визначення кольору може бути виконано на основі розгляду безлічі автомобілів.

¹³ Робота зі створення зв'язків послідовностей з іншими послідовностями дозволила в тілі нейронних мереж прискорити прогрес машинного перекладу.

обробляються прихованими шарами нейронної мережі. У мережах, які розпізнають мови або тексти початкові дані ще потрібно перетворити в векторну форму, зручну для роботи мережі.

Текстові дані і звуки мови спочатку класифікували за заздалегідь добре сформульованими ознаками (за рахунок лінійних класифікаторів), тобто вони повинні бути представлені в просторі ознак. Отримані таким чином (за допомогою унітарної кодування - т. зв. розріджене кодування) початкові вектори тексту, що враховують такі ознаки (в структурі вектора ознак наявність ознаки в певній позиції відзначається одиницею, а його відсутність - нулем) мають розмірність вельми значну - сотні тисяч і більше, а у відповідних їм векторів занурення, створених вже машиною (щільне кодування), розмірність виявляється на багато порядків менше. Тобто зазвичай вхідний шар (шар занурення, шар пошуку відповідності) таким чином виконує функцію виділення цих ознак і створення векторів занурення. Але це не означає, що приховані шари в режимі глибокого навчання не можуть зайнятися класифікацією (самостійно або, що проблематично, за попередньо отриманими правилами) даних від попередніх шарів. Якщо класифікаторів недостатньо, то використовують так звану попередню обробку документа замінюючи слова словниковими статтями - лемами з попередньо підготовлених баз даних цих статей.

Обговоримо нижче механізми класифікації текстів за різними системами ознак. Спочатку відбувається розбиття тексту на лексеми (tokenization) зазвичай це слова або групи слів. Для цього поділу користуються прогалинами, і знаками пунктуації. Для кожного слова леми - словникові статті, що пояснюють особливості застосування слова, розроблені зазвичай вручну. Ці леми об'єднані в словники, лексичні ресурси, що формують ознаки слів, до яких є доступ машині. Причому слід мати на увазі, що слова змінюють структуру і видозмінюються (флексії), що також в цих базах даних враховується одночасно. Національний інститут стандартів (National Institute of Standards) США розробив набір даних Mnist (60 тисяч зображень і їх міток, набір для розробки і тестування містить 10 тисяч зображень і міток), що описує цілі числа-цифри. Існують лексичні ресурси, зазвичай створені вручну, це надійніше і допомагає машині. Наприклад, WordNet - лексичний ресурс, зібраний вручну, пов'язує слова з їх сінсетом - семантичним знанням, - поняттям. FrameNet і VerbNet - лексичні ресурси, зібрані вручну для дієслів. Є Paraphrase (PPDB) - база даних парафраз. Тексти можна порівнювати на основі лічильників слів і порядку букв і слів (для англійської мови) в тексті. Тексти формують Bag-of-words - BOW, тобто «мішки» слів і відбувається їх оцінка (зважування) наприклад програмою TF-IDF за ознаками, за частотою появи (term frequency) по літерним биграмам (або по n-грамам). Це називають тематичною класифікацією - по мішку слів і мішку словесних биграмм. Для виключення неоднозначності передбачена оцінка по контексту, в мінімальній частини це вікно - найближче оточення слова. Це наслідок з дистрибутивної гіпотези мови: сенс слова завжди можна вивести з їх контексту. Розроблено biRNN - архітектура для узагальнення віконних ознак. Також існують проекти виду Universal Treebank Project - визначення частин мови по мітках POS (Part-of-speech tag), по положенню слова в реченні. Якщо слова представлені в векторній формі вже після процедури занурення, то відстань між словами визначаються з використанням однієї з відомих метрик. При навчанні з учителем, що характерно для глибоких (багатосарових) систем, використовують оптимізацію, яка складається в процедурах регуляризації - спробах уникнути багатозначності і заснована на перевазі помилок, які змінюються разом раз від разу рішенням (тобто перенавчання). Також намагаються забезпечити мінімізацію помилок (втрат) при зворотному поширенні помилки.

4 Висновок.

Успіхи в розробці нейронних мереж нового покоління, які були створені вже на сучасній технічній базі, викликали нову хвилю інтересу до створення теорій їх функціонування і навчання. Тим більше, що зросли технічні можливості, які дозволили спростити архітектури мереж і повернутися до архітектури перцептрона, тобто схемами прямого поширення. Хоча згодом більшість мереж глибокого навчання все ж складно було приписати до перцептронів, так як більш-менш загальноприйнята термінологія має на увазі називати перцептронами мережі з невеликою кількістю шарів і прямим поширенням сигналу, тоді як багато прикладних мереж вміщують в себе згорткові шари, dropout шари, pooling шари, зворотний зв'язок, шари підтримки, паралельні мережі). Так само не можна сказати, що саме мережі з прямим поширенням стали домінувати на ринку, швидше більшість завдань несли досить прикладний і простий характер,

рішення яких дозволяло грубо кажучи просто застосовувати багатошаровий перцептрон. Також відродився інтерес до навчання з учителем і використанню навчання зі зворотним поширенням помилки. Це надихнуло теоретиків на створення теорій так званого глибокого навчання, ідея якого полягала в послідовному навчанні шарів мережі, число яких вже можна було збільшувати.

Для мереж з декількома шарами в межах до десятка можна було використовувати пошарове навчання, хоча це вже було проблемою і вимагало використання технології паралельних обчислень і складного програмного забезпечення. Часто основною причиною відмови від спроб вирішувати будь-які проблеми простим збільшенням шарів, стала відсутність будь-яких альтернатив методом зворотного поширення помилки на базі градієнтного спуску, а саме тенденції загасання градієнта на якомусь проміжному шарі, що, хоч і вирішується певним чином, все таки не дозволяє додавати нескінченну кількість шарів і отримувати потрібні зміни ваг на найперших шарах. А ось для великих нейронних мереж пошарове навчання, яке розвивали теорії глибокого навчання по всій видимості застосовувати буде важко. Тому інтерес теоретиків в області глибокого навчання відволікся від нейромережових систем великої розмірності і в більшій мірі поширився на мережеві структури, які більше відповідали теоретичним уявленням. Але не можна не враховувати сучасні можливості технологій попереднього навчання. Існують проекти архітектур¹⁴, які під час навчання потребують мінімумі вихідних даних, які спираються на комбінацію навчених мережових рішень. Особливо для мереж, що використовують мову і тексти природної мови, на яких людство створило вражаючий монумент Знання. Крім того, практичне застосування можливостей сучасних нейронних систем безпосередньо пов'язано зі спілкуванням з людьми їхньою мовою.

ЛІТЕРАТУРА

1. Мак-Каллок У. С., Питтс В. Логическое исчисление идей, относящихся к нервной активности Архивная копия от 27 ноября 2007 на Wayback Machine // Автоматы / Под ред. К. Э. Шеннона и Дж. Маккарти. – М.: Изд-во иностр. лит., 1956. – С. 363–384. (Перевод английской статьи 1943 г.). <http://raai.org/library/books/mcculloch/mcculloch.pdf>
2. Hecht-Nielsen R. Kolmogorov's Mapping Neural Network Existence Theorem, IEEE First Annual Int. Conf. on Neural Networks, San Diego, 1987, Vol. 3, pp. 11–13. <https://cs.uwaterloo.ca/~y328yu/classics/Hecht-Nielsen.pdf>
3. Колмогоров А. Н. О представлении непрерывных функций нескольких переменных суперпозициями непрерывных функций меньшего числа переменных // Известия АН СССР, 108 (1956), с. 179–182
4. Lorentz George. Metric entropy, widths, and superpositions of functions // American Mathematical Monthly . 1962, vol. 69, P.469–485. <https://www.tandfonline.com/doi/abs/10.1080/00029890.1962.11989915>
5. Sprecher D. A. On the structure of continuous functions of several variables // Transactions of the American Mathematical Society 1965, vol. 115, P. 340–355. <https://www.jstor.org/stable/1994273>
6. Яновский В. В. Коллективный интеллект. НТК «Институт монокристаллов» НАН Украины – Киев: Наукова Думка, 2020.
7. Петров А. П. О возможностях перцептрона // Известия АН СССР, Техническая кибернетика. –1964. – №6.
8. Бонгард М. М. Проблемы узнавания. – М.: Физматгиз, 1967.
9. Круглов В. В., Борисов В. В. Искусственные нейронные сети. Теория и практика – 2 изд. – М. Горячая линия - Телеком, 2002 – 382с. <https://studizba.com/files/show/djvu/1761-1-kruglov-v-v-borisov-v-v-iskusstvennye.html>
10. Куклин В. М. Особенности развития искусственного интеллекта на современном этапе / Вісник Харківського національного університету імені В.Н. Каразіна, серія «Математичне моделювання. Інформаційні технології. Автоматизовані системи управління», 2018 с. 34-40 http://www.irbis-nbuv.gov.ua/cgi-bin/irbis_nbuv/cgiirbis_64.exe?I21DBN=LINK&P21DBN=UJRN&Z21ID=&S21REF=10&S21CNR=20&S21STN=1&S21FMT=ASP_meta&C21COM=S&2_S21P03=FILE=&2_S21STR=VKhIM_AM_2018_40_6

¹⁴ <https://towardsdatascience.com/neural-network-architectures-156e5bad51ba>

11. Куклін В. М. Подання знань і операції над ними; навчальний посібник. / В. М. Куклін. Х. : ХНУ імені В. Н. Каразіна, 2019. 164 с http://www-csd.univer.kharkov.ua/wp-content/uploads/2017/09/Kuklin_Podannya-znan....pdf.
12. Гушин І.В. Моделювання фізичних процесів із використанням технології CUDA : монографія / І. В. Гушин, В. М. Куклін, О. В. Мішин, О. В. Приймак. – Харків: ХНУ імені В. Н. Каразіна, 2017. – 116 с. <http://ekhnuir.univer.kharkov.ua/handle/123456789/14304>
13. Гудфеллоу Я., Бенджио И., Курвилль А. Глубокое обучение / пер. с англ. А. А. Слинкина. – 2-е изд., испр. – М.: ДМК Пресс, 2018. – 652 с.
14. Гольдберг Й. Г. Нейросетевые методы в обработке естественного языка / пер. с англ. А. А. Слинкина. – М.: ДМК Пресс, 2019. – 282 с. <https://www.livelib.ru/book/1003171275-nejrosetevye-metody-v-obrabotke-estestvennogo-yazyka-goldberg-j>

REFERENCES

1. W.S. McCulloch, W. Pitts. Logical calculus of ideas related to nervous activity Archived November 27, 2007 at the Wayback Machine // Ed. C.E. Shannon and J. McCarthy. - M. : Publishing house of foreign. lit., 1956. - pp. 363–384. (Translation of an English article of 1943). [in Russian]
2. Hecht-Nielsen R. Kolmogorov's Mapping Neural Network Existence Theorem, IEEE First Annual Int. Conf. on Neural Networks, San Diego, 1987, Vol. 3, pp. 11–13.
3. A.N. Kolmogorov On the representation of continuous functions of several variables by superpositions of continuous functions of a smaller number of variables // *Izvestiya AN SSSR*, 108 (1956), p. 179-182 [in Russian]
4. Lorentz George. Metric entropy, widths, and superpositions of functions // *American Mathematical Monthly* . 1962, vol. 69, P.469–485.
5. Sprecher D. A. On the structure of continuous functions of several variables // *Transactions of the American Mathematical Society* 1965, vol. 115, P. 340–355.
6. V.V. Yanovsky Collective intelligence. STC "Institute of Single Crystals" NAS of Ukraine - Kiev: *Naukova Dumka*, 2020 [in Russian]
7. A.P.Petrov On the possibilities of the perceptron // *Izvestiya AN SSSR, Technical cybernetics*. – 1964. - No. 6. [in Russian]
8. M.M. Bongard Problems of recognition. - M. : *Fizmatgiz*, 1967. [in Russian]
9. V.V. Kruglov , V.V. Borisov Artificial neural networks. Theory and Practice - 2nd ed. - M. Hot line - *Telecom*, 2002 - 382s [in Russian]
10. V.M. Kuklin Features of the development of artificial intelligence at the present stage / *Bulletin of the Kharkiv National University of V.N. Karazina, seria "Mathematical model. Information technologies. Automated control systems "*, 2018 p. 34-40 [in Russian]
11. V.M. Kuklin Submission of knowledge and operations over them; tutorial. / V. M. Kuklin. Kh.: KhNU imeni V.N. Karazin, 2019.164 p. [in Ukrainian]
12. I.V. Gushchin Modeling of physical processes using CUDA technology: monograph / I.V. Gushchin, V.M. Kuklin, O.V. Mishin, O.V. Priymak. - Kharkiv: VN Karazin KhNU, 2017. - 116 p [in Ukrainian]
13. J. Goodfellow ,I. Benjio , A. Courville. Deep learning / trans. from English A. A. Slinkina. - 2nd ed., Rev. - M. : DMK Press, 2018 .-- 652 p [in Russian]
14. Goldberg YG Neural network methods in natural language processing / per. from English A. A. Slinkina. - M. : DMK Press, 2019 .-- 282 p. [in Russian]