

Д. А. Іванов

ЗМЕНШЕННЯ ЧАСУ НА НАВЧАННЯ МОДЕЛЕЙ ЗА ДОПОМОГОЮ ТРАНСФЕРНОГО НАВЧАННЯ

У цій статті досліджується наскільки використання трансферного навчання може скоротити час навчання та ресурси, необхідні для навчання моделей глибокого навчання, зокрема на прикладі архітектури ResNet-50. Основна ідея трансферного навчання полягає у використанні вже навчених шарів моделі для виконання нових завдань, що дозволяє уникнути повного циклу навчання моделі з нуля. У процесі трансферного навчання певна кількість шарів моделі заморожуються, оскільки вони відповідають за розпізнавання загальних ознак, таких як контури або текстури, наприклад. Навчання відбувається тільки на останніх шарах моделі, які відповідають за виконання конкретного завдання. Це значно зменшує час та ресурси, що необхідні для навчання нових моделей. Результатами дослідження є визначення формул, що демонструють логарифмічне зростання витрат часу та ресурсів при використанні трансферного навчання, яке є менш інтенсивним порівняно з лінійним зростанням витрат часу та ресурсів під час класичного навчання моделі. Наведені формули показують, що для кожної нової моделі необхідно навчати лише 5 шарів, тоді як 45 шарів залишаються незмінними, що забезпечує значну оптимізацію. Ця кількість шарів використовується тільки при використанні архітектури ResNet-50, оскільки дослідження проводилось на цій архітектурі. Дослідження також доводить універсальність методу трансферного навчання, оскільки цей підхід може використовуватися і з іншими архітектурами. Різниця буде лише в тому, що за зміни архітектури, зміниться кількість шарів для навчання загальних ознак, та кількість шарів, що потрібні для спеціалізованих завдань, а також середній час навчання одного шару. Попри це, навчання моделей з іншими архітектурами за допомогою трансферного навчання буде використовувати менше ресурсів та часу для процесу навчання.

Ключові слова: трансферне навчання, ResNet-50, оптимізація часу, глибоке навчання, машинне навчання, логарифмічне зростання, комп'ютерний зір, скорочення ресурсів.

Вступ

З розвитком технологій штучного інтелекту (ШІ) та машинного навчання (ML) підвищуються вимоги до ефективності та швидкості навчання моделей. У багатьох сучасних застосуваннях, таких як медична діагностика, автономні транспортні засоби та системи відеоспостереження, моделі мають навчатися на величезних обсягах даних за короткий час. Класичні методи навчання моделей з нуля часто потребують великих обчислювальних ресурсів і часу, що обмежує їхню ефективність в реальних умовах.

Трансферне навчання стало одним із найефективніших підходів для вирішення цих проблем, оскільки дозволяє використовувати вже навчені моделі для швидкої адаптації до нових завдань, що значно скорочує час і ресурси, необхідні для навчання. Особливо ефективним цей метод є для моделей, які використовуються у задачах комп'ютерного зору, де велика кількість базових ознак, таких як контури, форми та текстури, може бути перенесена між різними наборами даних і завданнями.

У цій науковій роботі відбувається визначення наскільки використання трансферного навчання пришвидшує та зменшує використання ресурсів. В рамках дослідження розглядається архітектура ResNet-50 як приклад моделі глибокого навчання, яка широко використовується в задачах комп'ютерного зору.

Метою статті є визначення зменшення часу та ресурсів на навчання моделей при використанні трансферного навчання. **Мета** включає конкретне завдання – визначення наскільки зменшиться час при використанні трансферного навчання та довести це за допомогою математичних формул. За допомогою цих формул можна визначити ефективність цього методу.

Постановка проблеми

У сучасних задачах машинного навчання, зокрема у розпізнаванні зображень і роботі з великими обсягами даних, навчання моделей стає вкрай ресурсомістким процесом, що потребує значного часу та обчислювальних потужностей. В умовах швидкого розвитку технологій і підвищених вимог до обробки даних у реальному часі, класичне навчання моделей з нуля є неефективним рішенням, особливо в ситуаціях, де ресурси обмежені або необхідна швидка адаптація моделі до нових завдань.

Трансферне навчання пропонує альтернативний підхід, який дозволяє використовувати вже навчені моделі, адаптуючи їх до нових задач за менший час і з меншими ресурсними затратами. Проте залишається питання, наскільки ефективно цей підхід може скоротити витрати часу під час навчання моделей у різних сферах. Особливо це важливо для задач, де є потреба в оперативній обробці інформації або обмеженість у доступі до великих обсягів даних для навчання з нуля.

Таким чином, актуальним є завдання дослідження покращення процесу навчання за допомогою методу трансферного навчання, зокрема на прикладі моделі ResNet-50, що широко використовуються у комп'ютерному зорі. Це дозволить оцінити, як саме скорочуються часові витрати та наскільки покращується продуктивність моделей при використанні попередньо навчених шарів для нових завдань, порівняно з класичними методами навчання.

Аналіз останніх досліджень і публікацій

Pan і Yang [1] детально описали стратегії трансферного навчання, які значно скорочують час навчання моделей. Вони показали, що перенесення знань між схожими завданнями зменшує потребу в великих обсягах даних і обчислювальних ресурсах. Однак для суттєво відмінних завдань цей підхід залишається менш ефективним, що потребує додаткових досліджень.

Weiss, Khoshgoftaar та Wang [2] зосередили увагу на налаштуванні тонкої та вилученні ознак, які дозволяють адаптувати попередньо навчені моделі. Цей підхід пришвидшує навчання нових моделей, але проблема оптимального вибору параметрів для різних завдань залишається невирішеною.

Pratt і Mostow [3] розглянули можливість передачі знань між нейронними мережами для прискорення навчання. Вони продемонстрували, що для схожих завдань такий підхід працює, але складнощі виникають при застосуванні до завдань, які кардинально відрізняються за характеристиками.

Torrey і Shavlik [4] дослідили практичні застосування попередньо навчених моделей. Вони підкреслили, що використання таких моделей значно скорочує час навчання, але не всі моделі однаково добре адаптуються до нових завдань, що обмежує швидкість навчання в деяких сценаріях.

Визначення трансферного навчання

Трансферне навчання представляє собою методику в галузі машинного навчання, яка сприяє ефективному перенесенню знань, отриманих з одного завдання, для вирішення іншого, часто не пов'язаного завдання. Цей підхід має значний потенціал для зниження часу та витрат, пов'язаних з розробкою нових моделей навчання з нуля [1].

Розглянемо приклад: модель, розроблена для ідентифікації котів на фотографіях, може бути адаптована для визначення собак. Цей процес не лише підкреслює здатність моделі до адаптації, але й демонструє універсальність властивостей, які можуть бути використані у широкому спектрі завдань [2].

Принцип роботи трансферного навчання

Принцип роботи трансферного навчання полягає у використанні попередньо навчених моделей як відправної точки для навчання моделей для нових завдань. Таким чином, модель, навчена на одному завданні (джерелом), може використовувати свої знання для вирішення іншого завдання (цілі). Це можливе завдяки тому, що багато завдань мають спільні риси, які можуть бути корисними для вирішення нових проблем. Наприклад, модель, яка навчилася розпізнавати загальні шаблони в зображеннях, такі як контури або текстури, може успішно використовувати ці знання для розпізнавання об'єктів у нових, раніше невідомих наборах даних [2]. Такий підхід дозволяє суттєво скоротити час і ресурси на навчання нових моделей.

Процес трансферного навчання складається з декількох етапів, кожен з яких відіграє важливу роль. Перший етап – початкове навчання моделі. На цьому етапі модель навчається на великому обсязі даних, що називається "джерело". Це можуть бути загальні або конкретні набори даних залежно від галузі застосування. Модель вивчає базові шаблони, такі як форми, текстури, контури, зв'язки слів або синтаксичні структури, залежно від типу даних. Pan і Yang вказують, що цей етап є критичним для побудови "базу знань" для моделі [1, 9].

Також на цьому етапі модель проходить стандартний процес навчання, використовуючи методи глибокого навчання, такі як зворотне поширення помилки (backpropagation) і градієнтний спуск. Завдання моделі полягає в тому, щоб навчитися узагальнювати певні характеристики з навчального набору даних, досягаючи максимальної точності в розпізнаванні цих характеристик. Це створює "базу знань", яку модель буде використовувати для нових завдань [1].

Наступний етап – це етап заморожування загальних шарів. На цьому етапі частина моделі, яка вже навчилася розпізнавати загальні риси (наприклад, контури або текстури), залишається незмінною. Це називається "заморожуванням" шарів. Вони більше не оновлюються під час подальшого навчання, оскільки вже володіють знаннями, які можуть бути корисними для багатьох завдань [8].

На третьому етапі модель адаптується до нового завдання, використовуючи знання, отримані на першому етапі. Якщо модель спочатку навчилася розпізнавати певні об'єкти на великому наборі даних, вона може бути адаптована до нового. Тут модель не вчиться з нуля, вона просто переналаштовується на нове завдання, використовуючи попередні знання. Цей етап економить час і ресурси, адже модель вже володіє базовими знаннями про об'єкти [2].

На останньому етапі модель проходить через додаткове налаштування і оптимізацію, щоб досягти максимальної продуктивності на нових даних. Це може включати налаштування гіперпараметрів, таких як швидкість навчання або кількість епох навчання. Цей процес може вимагати кількох ітерацій, щоб досягти найкращих результатів, що підвищує точність і продуктивність моделі [8].

Переваги трансферного навчання

Трансферне навчання пропонує переваги, які можуть революціонізувати процес розробки та впровадження моделей. Цей метод спрощує використання великих і складних моделей, оскільки дозволяє адаптувати вже навчені моделі до нових задач, тим самим знижуючи витрати та час на розробку та навчання моделей з нуля. Застосування трансферного навчання сприяє більшій продуктивності та ефективності в багатьох областях, від комп'ютерного зору до обробки природної мови [2].

Однією з ключових переваг трансферного навчання є зменшення часу, необхідного для навчання моделей. Використовуючи попередньо навчені моделі, ці моделі швидше адаптуються до нових завдань, що дозволяє реалізувати рішення швидше, ніж при навчанні з нуля. Наприклад, в дослідженні, проведеному Weiss і колегами, було виявлено, що моделі, які використовували трансферне навчання, досягали порівняльної точності значно швидше, ніж моделі, навчені з нуля [2].

Наступною важливою перевагою є покращення продуктивності моделей на нових

завданнях. Трансферне навчання дозволяє моделям використовувати загальні властивості та знання, здобуті під час навчання на одних даних, для покращення результатів на інших. Це особливо корисно у задачах, де дані обмежені або важкодоступні, дозволяючи створювати ефективніші та точніші моделі [1].

Зниження потреби у великих наборах даних є ще однією суттєвою перевагою. В багатьох випадках, збір великої кількості анотованих даних може бути вкрай дорогим або неможливим з етичних причин, особливо у таких сферах як медицина. Трансферне навчання може мінімізувати цю потребу, використовуючи вже наявні набори даних для вивчення основних властивостей, що веде до зменшення загальних витрат на розробку моделей [3].

Застосування трансферного навчання також сприяє більш ефективному використанню обчислювальних ресурсів. Оскільки моделі потребують меншого часу навчання, це знижує навантаження на обчислювальні системи і зменшує вартість використання обчислювальних ресурсів [4].

Однією з останніх, але не менш важливих переваг трансферного навчання є покращена здатність моделей узагальнювати з одного домену на інший. Моделі, які розробляються з використанням трансферного навчання, часто краще адаптуються до нових умов, оскільки вони вже "навчилися" виявляти корисні особливості в одному контексті, які можуть бути застосовні й у іншому [5].

Вибір моделі для дослідження трансферного навчання

Для дослідження використано модель глибокого навчання ResNet-50, яка є типовим вибором для завдань, пов'язаних з комп'ютерним зором [6]. Цю модель було обрано, оскільки вона є однією з популярних і широко використовуваних моделей. **Метою дослідження** є визначити, наскільки зменшиться час для навчання декількох моделей при використанні трансферного навчання порівняно з класичним навчанням з нуля.

Архітектура ResNet-50 та її застосування у трансферному навчанні

Модель ResNet-50 складається з 50-ти шарів, що включають конволюційні шари, pooling шари та залишкові блоки (residual blocks), які використовують пропускні з'єднання (skip connections) для обробки ознак без втрати інформації [7]. Це допомагає уникнути проблеми затухання градієнтів, характерної для дуже глибоких мереж. Початкові шари відповідають за розпізнавання базових ознак, таких як контури, форми, і текстури.

У процесі трансферного навчання 45 початкових шарів заморожуються, оскільки вони вже здатні розпізнавати загальні ознаки. Лише останні 5 шарів залишаються активними для перенавчання на нових даних, що дозволяє моделі адаптуватися до специфічних задач. Така архітектура значно скорочує час навчання, оскільки більшість шарів не потребують повторної обробки.

Час навчання повної моделі ResNet-50 може тривати кілька днів на стандартних графічних процесорах, залежно від обсягу даних та обладнання. У випадку трансферного навчання, де потрібно навчати лише 5 шарів, цей час суттєво скорочується, іноді до кількох годин [10].

Дослідження трансферного навчання

Одна з головних переваг трансферного навчання – це скорочення витрат часу та ресурсів на процес навчання моделей. Як вже було сказано, це ефективно, коли відсутня велика кількість даних для навчання, і модель може використати як основу модель, яка розпізнає загальні ознаки, або коли завдання моделей доволі схожі, і можна використати базову модель і не вчити всі моделі з нуля.

Як було розглянуто вище, для навчання модель ResNet-50 використовує 50 шарів, і якщо потрібно навчати декілька моделей на завдання, що схожі, то потрібно кожен раз використовувати 50 шарів для навчання. Цю закономірність можна описати формулою:

$$N_{\text{класичне}} = n_{50} * k \quad (1)$$

де k – це кількість моделей для навчання, n_{50} – кількість шарів для навчання моделі ResNet-50.

При використанні трансферного навчання, формула для обрахунку кількості шарів, що потрібні для навчання моделей, буде інша. Оскільки при використанні трансферного навчання, нові моделі будуть використовувати як основу, модель що має 45 навчених шарів та розпізнає загальні ознаки, їм потрібно буде навчати лише 5 шарів. Цю закономірність можна описати формулою:

$$N_{\text{трансферне}} = n_{45} + (n_5 * k) \quad (2)$$

де k – це кількість моделей для навчання, n_{45} – кількість шарів для навчання моделі ResNet-50 на початковому етапі, n_5 – кількість шарів для навчання нової моделі ResNet-50.

За допомогою формул 1 та 2, можна визначити час, який потрібен на навчання моделей. Закономірність часу класичного навчання можна описати наступною формулою:

$$T_{\text{класичне}} = N_{\text{класичне}} * t \quad (3)$$

де $N_{\text{класичне}}$ – це кількість шарів для навчання, t – середній час навчання одного шару.

Для трансферного навчання, де заморожено 45 шарів, закономірність часу можна описати наступною формулою, яка відображає зниження витрат ресурсів:

$$T_{\text{трансферне}} = N_{\text{трансферне}} * t \quad (4)$$

де $T_{\text{трансферне}}$ – це кількість шарів для навчання, t – середній час навчання одного шару.

Закономірність витрат часу можна виразити через логарифмічну форму для порівняння масштабів. Спочатку виражено закономірність для класичного навчання:

$$\log(T_{\text{класичне}}) = \log(n_{50}) + \log(k) + \log(t) \quad (5)$$

Формула (5) показує, що час навчання зростає лінійно зі збільшенням кількості моделей k . Кожна нова модель додає до загального часу навчання 50 шарів, що відображає прямолінійний зв'язок між кількістю моделей і часом.

Закономірність витрат часу для трансферного навчання також можна виразити через логарифмічну форму. Вона виглядає так:

$$\log(T_{\text{трансферне}}) = \log(n_{45} + n_5 * k) + \log(t) \quad (6)$$

Формула (6) показує логарифмічне зростання. Це означає, що збільшення витрат часу відбувається повільніше з кожним наступним приростом порівняно з лінійним зростанням. У формулі (6) це зростання пов'язане з тим, що лише частина моделі перенавчається (5 шарів), тоді як інші шари (45) залишаються замороженими. Логарифмічне зростання демонструє, що при збільшенні кількості моделей k , час навчання збільшується менш інтенсивно порівняно з класичним підходом, де всі шари навчаються заново.

Висновки

У цій науковій статті визначено, наскільки трансферне навчання зменшує витрати часу та ресурсів порівняно з класичним підходом. Аналіз ResNet-50 показав, що використання заморожених шарів дозволяє суттєво скоротити час навчання без зниження точності, що підтверджується формулою (6). Якщо взяти середній час навчання одного шару як 1 година і розрахувати час, наприклад, для трьох моделей, то виходить, що для навчання з нуля потрібно буде приблизно 150 годин, а для трансферного навчання – 60 години. Це

демонструє, що трансферне навчання показує логарифмічне зростання витрат часу, що є повільнішим за лінійне зростання при класичному підході. Це дозволяє масштабувати навчання із меншими часовими та ресурсними витратами, що є критично важливим для великих задач.

Результати цього дослідження є універсальними, оскільки при зміні моделі, закономірність витрат часу визначатиметься за тією ж формулою. Єдина зміна полягає в тому, що при зміні моделі буде інша кількість шарів для навчання моделі, і відповідно буде інша кількість шарів для навчання на початковому етапі.

Таким чином, результати дослідження показують, що трансферне навчання дозволяє зменшити витрати часу та ресурсів, що робить його практичним інструментом для застосувань, особливо в галузях з обмеженими ресурсами або потребою швидкої адаптації моделей.

СПИСОК ЛІТЕРАТУРИ

1. Pan S. J. A Survey on Transfer Learning / S. J. Pan, Q. Yang // IEEE Transactions on Knowledge and Data Engineering. – 2010. – № 22 (10). – P. 1345 – 1359.
2. Weiss K. A Survey of Transfer Learning / K. Weiss, T. M. Khoshgoftaar, D. Wang // Journal of Big Data. – 2016. – № 3. – P. 9.
3. Pratt L. Y. Direct Transfer of Learned Information Among Neural Networks / L. Y. Pratt, J. Mostow // Proceedings of the Ninth National Conference on Artificial Intelligence (AAAI-91), Anaheim, CA, USA, July 14-19, 1991. – P. 584 – 589.
4. Torrey L. Transfer Learning / L. Torrey, J. Shavlik // Handbook of Research on Machine Learning Applications and Trends / eds. E. Soria, J. Martin, R. Magdalena, M. Martinez, A. Serrano. – IGI Global, 2009. – P. 242 – 264.
5. How Transferable Are Features in Deep Neural Networks? / J. Yosinski, J. Clune, Y. Bengio [et al.] // Advances in Neural Information Processing Systems. – 2014. – № 27. – P. 3320 – 3328.
6. Deep Residual Learning for Image Recognition / K. He, X. Zhang, S. Ren [et al.] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2016. – P. 770 – 778.
7. Learning and Transferring Mid-level Image Representations Using Convolutional Neural Networks / M. Oquab, L. Bottou, I. Laptev [et al.] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2014. – P. 1717 – 1724.
8. A Comprehensive Survey on Transfer Learning / F. Zhuang, Z. Qi, K. Duan [et al.] // Proceedings of the IEEE. – 2020. – Vol. 109, Issue 1. – P. 43 – 76.
9. A Survey on Deep Transfer Learning / C. Tan, F. Sun, T. Kong [et al.] // The 27th International Conference on Artificial Neural Networks (ICANN), Rhodes, Greece. – 2018. – C. 270 – 279.
10. The Basics of ResNet-50 [Electronic resource] / I. Mostafa // Weights & Biases Reports. – 2023. – Access mode: <https://wandb.ai/mostafaibrahim17/ml-articles/reports/The-Basics-of-ResNet50---Vmlldzo2NDkwNDE2>.

Стаття надійшла до редакції 26.08.2024.

Стаття пройшла рецензування 20.09.2024.

Іванов Дмитро Анатолійович – аспірант кафедри інженерії програмного забезпечення, e-mail: www.ivanovda@gmail.com, <https://orcid.org/0000-0002-7386-4497>.

Державний Університет “Житомирська політехніка”.