

УДК 004.6

ДОСЛІДЖЕННЯ МЕТОДІВ ОДНОМОМЕНТНЕ НАВЧАННЯ ДЛЯ МУЛЬТИМОДАЛЬНИХ ДАНИХ

Стахевич А.В.

Науковий керівник – к.т.н., доц. Турута О.П.

Харківський національний університет радіоелектроніки, каф. ШІ
м. Харків, Україна

тел.: +38(099) 982-49-46, e-mail: anna.stakhevych@nure.ua

This work is devoted to the study of one-shot learning methods for generating image captions in the Ukrainian language with a very small set of training data.

В сучасному світі дані відіграють дедалі важливішу роль у різних галузях і секторах, технології комп'ютерного зору та обробки зображень стають невід'ємною частиною нашого життя. У цьому контексті опис зображень, який дозволяє комп'ютерам аналізувати і розуміти зміст зображень так само, як це робить людина, є одним з основних напрямків розвитку цієї технології. У зв'язку з цим напрямком з опису зображень стають все більш затребувані серед фахівців, які хочуть поглибити свої знання та удосконалити професійні навички в галузі комп'ютерного зору. Особливо ця тема актуальна для реалізації моделі, яка може генерувати підписи українською мовою.

Для створення надійної моделі, яка може описати детально зображення необхідна велика кількість надійних прикладів. Інколи це стає проблемою, оскільки не завжди є можливість отримати достатній набір даних. У цьому випадку необхідно розробляти більш надійні системи, які можуть навчатися не гірше інших моделей на маленькому наборі даних, та видавати більш точні результати. Існують багато різних технологій, які вже зарекомендували себе у даному напрямку. Однією з таких стала технологія одномоментного навчання. Така технологія дає змогу навчити модель розпізнавати та описувати зображення при невеликій кількості прикладів. Це можна порівняти з інтуїтивним навчанням у людини, де людині показують декілька прикладів, а після цього вони можуть описати на основі попереднього досвіду.

Одномоментне навчання – це тип машинного навчання, в якому моделі навчаються розпізнавати новий клас об'єктів, використовуючи лише один або кілька прикладів. Традиційне машинне навчання вимагає наявності великого набору даних, щоб навчити модель розпізнавати новий об'єкт, тоді як одномоментне навчання може навчити модель на невеликій кількості прикладів.

Одним із варіантів реалізації цього підходу є система, яка створює підписи до зображення. Підхід розв'язання цієї задачі складається з використання архітектури нейронної мережі, яка поєднує візуальний кодер

і декодер мови. Візуальний кодер зазвичай є згортковою нейронною мережею (CNN), яка вилучає характеристики з вхідного зображення, тоді як мовний декодер зазвичай є рекурентною нейронною мережею (RNN) або її варіація, яка генерує вихідний підпис.

Одним із обмежень для генерування підписів до зображень є те, що вони можуть бути не такими ефективними при обробці варіацій у зображень чи стилі підписів, як інші методи одномоментного навчання, оскільки вони покладаються на фіксоване відображення між входами та виходами. Однак невелика кількість еталонних прикладів може бути дуже ефективною для створення точного і зрозумілого контексту.

Інший підхід полягає у використанні мультимодального механізму уваги, який навчається звертати увагу на різні частини зображення та генерувати відповідні слова в підписі. У цьому підході мережа кодувальника створює набір ознак, які використовуються механізмом уваги для вибіркового звернення до різних частин зображення під час генерації підпису. Механізм уваги також тренується за допомогою підходу одномоментного навчання, коли він точно налаштовується на невеликому наборі контрольних прикладів.

Ще одним варіантом реалізації може бути передавальне навчання. Головна ідея цього методу полягає у тому, що на початку ми вже маємо нейронну мережу, яка була навчена на великій кількості даних для виконання певної задачі. Цю модель можна точно налаштувати на меншому наборі нових зображень і підписів, які можуть мати інший розподіл або словниковий запас порівняно з вихідним набором даних. Це дозволяє моделі навчатися на нових даних, одночасно використовуючи знання, отримані з вихідного набору даних, створюючи точнішу та ефективнішу модель.

Трансферне навчання особливо корисно, коли нове завдання має обмежені дані, оскільки воно може допомогти подолати проблему надмірного оснащення та покращити узагальнення моделі. Це також корисний прийом для зменшення кількості часу та ресурсів, необхідних для навчання нової моделі з нуля.

Наступний підхід є навчання з підкріпленням. Його можна використовувати для створення підписів на основі системи винагород. Модель генеруватиме заголовок, а винагорода залежатиме від того, наскільки точний заголовок. Потім модель регулювала б свої параметри на основі винагороди, щоб покращити свої можливості субтитрів.

Загалом мета цих методів полягає в тому, щоб створити модель, яка може генерувати підпис для нового зображення на основі невеликого набору прикладів, не вимагаючи великих обсягів анотованих навчальних даних.