



КОМП'ЮТЕРНА ІНЖЕНЕРІЯ ТА ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ

УДК 005.8:316.422

DOI <https://doi.org/10.17721/ISTS.2020.4.58-62>Б. Бондаренко, orcid.org/0000-0001-6431-3434,
daren.24@ukr.netЮ. Я. Самохвалов, orcid.org/0000-0001-5123-1288,
yu131053@gmail.com

Київський національний університет імені Тараса Шевченка, Київ, Україна

ПОШУК МУЛЬТИМЕДІЙНОЇ ІНФОРМАЦІЇ НА ОСНОВІ НЕЙРОННИХ МЕРЕЖ

Розглянуто підходи до використання нейронних мереж для пошуку мультимедійної інформації. Розроблення методів пошуку мультимедійної інформації необхідне через велику кількість такої інформації. Традиційні методи пошуку мультимедійної інформації мають високу швидкість оброблення даних, але низьку точність через відсутність можливості виконання семантичного пошуку. Використання нейронних мереж дозволяє здійснювати семантичний пошук, що збільшує його точність і повноту. Наведено підходи використання нейронних мереж на етапах індексування та пошуку мультимедійної інформації. За допомогою нейронної мережі аналізують мультимедійний файл і виконують його класифікацію. Результат класифікації файлу використовують для створення його текстового опису – анотації, яку порівнюють із запитом для визначення релевантності. Існує багато готових мереж для класифікації, використання яких дозволяє пришвидшити процес створення системи пошуку мультимедійної інформації, але неможливо створити нейронну мережу для класифікації всіх об'єктів реального світу, тому потрібно застосовувати декілька нейромереж. Також за допомогою нейронних мереж будують вектори ознак для мультимедійного файлу та пошукового запиту. Прості функції подібності, такі як косинус подібності, застосовують до побудованих векторів, для визначення семантичної близькості запиту та мультимедійного файлу. Причому пошуковий запит може бути як у текстовій формі, так і у вигляді будь-якого формату пошуку в мультимедійному файлі. Цей підхід дозволяє будувати оптимальну нейронну мережу під конкретну задачу. Нейронні мережі застосовують для порівняння побудованої анотації файлу та запиту, що підвищує точність і повноту пошуку, порівняно з традиційними методами, за рахунок здатності нейромереж враховувати семантичне значення тексту.

Ключові слова: нейронні мережі; пошук мультимедійної інформації; семантичний пошук; системи пошуку інформації; анотація даних.

1. ВСТУП

З поширенням інтернету з'явилась можливість передачі та зберігання великих обсягів інформації, зокрема мультимедійної. Це викликало потребу в розробленні методів пошуку інформації мультимедіафайлів. Існуючі підходи до пошуку текстової інформації не можливо застосовувати для пошуку мультимедійних даних. Тому створення ефективних методів пошуку мультимедійної інформації є важливим напрямом досліджень. Наявні методи, такі як аналіз гістограм кольорів зображення або частоти звуку [1], мають високу швидкість оброблення, але не здатні

виконувати семантичний пошук. Використання нейронних мереж дозволяє обробляти мультимедіафайли та виконувати семантичний пошук, що підвищує точність і повноту пошуку.

2. АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

Питання пошуку мультимедійної інформації висвітлено в джерелах [1–4]. Вони розглядалися відомими компаніями, такими як, Google, Apple та Amazon, які займаються збиранням і зберіганням даних. Указаній тематиці присвячена велика кількість міжнародних конференцій, се-

© Бондаренко Б., Самохвалов Ю. Я., 2020



мінарів і круглих столів. Проте питання пошуку мультимедійної інформації на основі використання нейромереж ще не достатньо повно висвітлено у працях вітчизняних і зарубіжних вчених. Тому застосування нейромережного підходу у пошуку мультимедійної інформації є актуальним напрямом досліджень.

3. ОСНОВНИЙ ТЕКСТ

Інформація в базі даних, з якою працює інформаційно-пошукова система, може бути структурованою або неструктурованою. Саме структура даних вимагає відповідного підходу до отримання корисних знань з інформаційних масивів.

Структуровані дані – це інформація, упорядкована й організована певним чином із метою її ефективного оброблення. [3]. Неструктуровані дані – це інформація, яка або не має наперед визначеної структури, або не організована в установленому порядку. Підвидом неструктурованих даних є мультимедійні, до яких належать зображення, аудіо- та відеофайли. Такі дані зберігають у первісному вигляді й аналізують за допомогою спеціальних методів.

Система пошуку інформації повинна мати можливість представляти та зберігати інформацію таким чином, щоб забезпечити її швидкий пошук [2]. Вона має бути здатною працювати з різними типами інформації, зокрема й мультимедійною. Щоб здійснювати пошук мультимедійних даних, система, як правило, повинна визначати їхні особливості, наприклад різноманітність кольорів і текстуру зображення або ритмічність аудіо. На рис. 1 показано загальну структуру системи пошуку інформації.

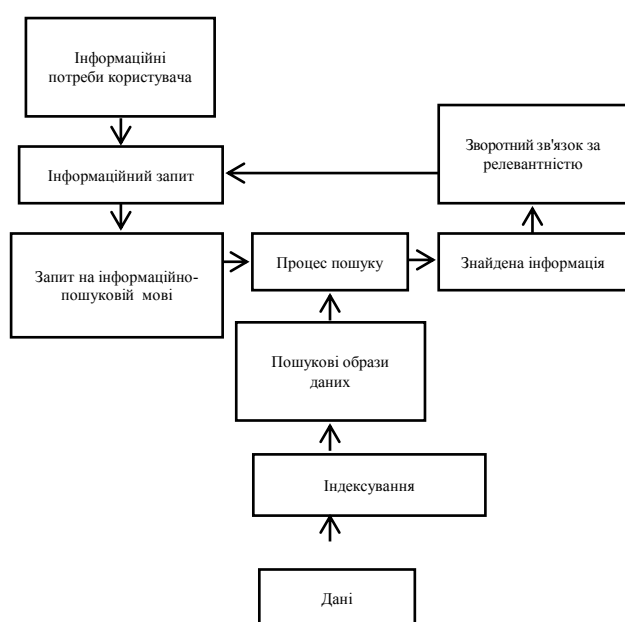


Рис. 1. Загальна структура системи пошуку інформації

Основними етапами роботи системи пошуку інформації є: індексування документів і створення їхнього пошукового образу (ПО); перетворення запиту користувача, сформульованого на природній мові, на запит мовою інформаційно-пошукової системи; пошук інформації і надання її користувачу.

Серед цих етапів індексування й пошук інформації є ключовими, тому що повинні враховувати семантичне значення документів. Тому використання на цих етапах нейронних мереж дозволить підвищити точність і повноту пошуку.

3.1. ІНДЕКСУВАННЯ МУЛЬТИМЕДІЙНОЇ ІНФОРМАЦІЇ

Щоб мати можливість швидкого пошуку інформації, системі потрібно побудувати її пошуковий образ. Для цього використовується індекс. Індексация дозволяє структурувати дані так, щоб була можливість пошуку документів за одним або кількома критеріями подібності [5]. У процесі індексації створюється пошуковий образ файлу, який потім порівнюється із запитом для визначення його релевантності. Для проведення індексації мультимедійної інформації розроблено різні методи. Наприклад, використання гістограм кольорів для оброблення зображень і використання прихованої марковської моделі для аудіофайлів [1, 2]. Одним із методів індексації є створення анотації – текстового опису файлу. Анотування файлу – це процес присвоєння йому ключового слова або списку ключових слів, що описує семантичне значення його вмісту.

Анотування зображень. Під час оброблення зображень і відео (як набору зображень) анотація може виконуватися на двох рівнях: локальному та глобальному. На локальному рівні зображення розглядається як сукупність об'єктів. Така анотація має на меті описати кожен об'єкт на зображенні у вигляді ключового слова або списку ключових слів. Глобальний рівень описує весь образ зображення і створює список ключових слів для опису його загального змісту.

Об'єкти, що мають однакоє семантичне значення, можуть мати різний вигляд на фото [4]. Анотування зображень традиційними методами є малоефективним, адже вони використовують прості ознаки зображення (наприклад, форма, гістограми кольорів і текстура), які обчислюються на основі характеристик пікселів, таких як, положення та колір. Під час створення анотації зображень за допомогою нейронних мереж класифікують об'єкти та явища, що зображені на них, та визначають пов'язані з ними ключові слова.

Існують різні архітектури нейронних мереж для класифікації. Майже всі вони використовую-



ють як основу згорткові шари та шари агрегування. За допомогою спеціально натренованих згорткових нейромереж можна виконувати класифікацію та опис зображень. Після класифікації зображення створюється його текстовий опис. У подальшому виконується семантичний пошук по описах зображень і знаходяться не лише конкретні об'єкти із зображень, а близькі до них за значенням. Наприклад, у разі пошуку слова *стілець*, система знаходить усі зображення на яких є *стілці*, а також об'єкти зі схожим значенням, наприклад *стіл*. Прикладом мережі класифікації зображень є Inception-ResNet-v2 – це згорткова нейронна мережа, яка натренована на більш ніж мільйони зображень із бази даних ImageNet [6]. Її архітектуру наведено на рис. 2. Мережа складається зі 164 шарів і може класифікувати зображення за 1000 категоріями різних предметів і тварин. Як результат, мережа отримала широкі можливості для класифікації зображень [6].

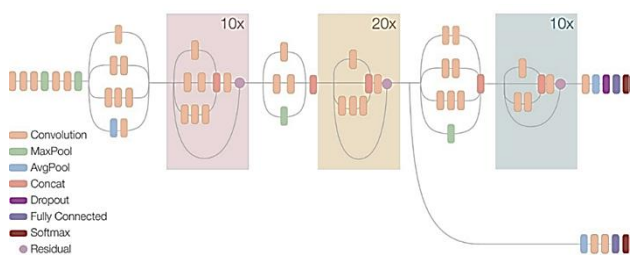


Рис. 2. Архітектура мережі для класифікації зображень

Мережа Inception-ResNet-v2 приймає на вхід зображення розміром 299 на 299 пікселів та визначає, що зображено на ньому. Її глибока архітектура складається з більше 30 згорткових шарів, що дозволяє показувати високу точність класифікації, але потребує значних затрат ресурсів на оброблення одного зображення.

Окрім класифікації, нейронні мережі можуть виконувати сегментацію, тобто визначати розмір і положення об'єкта на зображенні. Результат сегментації також може бути корисним при побудові анотації. Існуючі нейронні мережі не можуть класифікувати будь-які об'єкти, тому доцільно використовувати декілька нейромереж, натренованих для класифікації різних типів об'єктів. Створення універсальної системи класифікації – це ще один напрям для розвитку.

Анотування аудіоінформації. Пошук певного звуку або типу звуку (наприклад, промови певної людини або музики) може бути непростим завданням [7]. Немає стандарту класифікації звуків: два користувачі дадуть різний опис для одного звуку.

Для створення анотації аудіофайлів потрібно використовувати різні типи нейронних мереж,

адже кожен аудіофайл несе своє значення – якщо це запис промови, то потрібно розпізнати слова, якщо це музика, то можливо визначити її жанр. Часто в основі мереж для розпізнавання звуків лежать рекурентні нейронні мережі, адже вони здатні працювати з послідовностями даних, розподіленими в часі. Їх використовують для розпізнавання природної мови та прогнозування даних. Рекурентна нейронна мережа на виході додатково виводить прихований стан, який додається до наступних даних, та знову подається на вхід цієї ж мережі.

Різновидом рекурентних мереж є вентильний рекурентний вузол (Gated Recurrent Unit, GRU), який працює таким чином [8]. На вхід подають значення вхідного вектора x_t та вихідне значення попереднього блоку h_{t-1} . За ними розраховують значення векторів вузлів уточнення z_t та скидання r_t :

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z), \quad (1)$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r), \quad (2)$$

де W та U – матриці ваг, а b – вектор параметрів.

Значення вузла уточнення z_t допомагає моделі визначити, яку кількість минулої інформації (з попередніх етапів часу) потрібно передати далі. А вузла скидання r_t – яку кількість минулої інформації забути (не враховувати). Ці два вектори використовують для розрахунку прихованого стану даної ітерації мережі. Вихідні значення даного блоку h_t розраховують за формулою

$$h_t = z_t \circ h_{(t-1)} + (1 - z_t) \circ \tanh\left(\frac{W x_t + r_t \circ U h_{(t-1)} + b_h}{2}\right), \quad (3)$$

де $\circ$ – добуток Адамара.

Якщо значення елемента вектора r близьке до 1, то більшість інформації з попередньої ітерації не буде враховано. Причому, чим ближче значення z_t до 1, тим більше інформації з попереднього виходу мережі буде передано на наступну ітерацію.

Аудіофайл поділяють на частини однакової довжини й аналізують за допомогою рекурентних нейромереж. Це дозволяє виконати розпізнавання мови, записаної на аудіофайлі, або класифікувати його жанр, якщо це музика. Розпізнана інформація файлу використовується для формування його анотації.

Побудова вектора ознак. Для пошуку мультимедійної інформації, за допомогою нейронних мереж, будують вектор ознак медіафайла. Використовуючи мережу зі спеціально побудованою архітектурою, мультимедіафайл перетворюють на n -вимірний вектор, створюючи його пошуковий образ [3]. На рис. 4 зображено загальну архітектуру нейронної мережі перетворення зображення на вектор ознак. Перші шари таких мереж зазвичай є шарами згортки й агрегування, які потрібні для зменшення розмірності вхідного файлу. Після цих

шарів розташовують повнозв'язний шар, вихідне значення елементів якого стає вектором ознак. Для покращення точності пошуку, між шарами згортки та повнозв'язним шаром доцільно розташовувати декілька прихованих шарів.

Побудований вектор ознак для мультимедіа файла зберігає в собі особливості даних, наприклад, для зображення – це різноманіття кольорів, пропорції фігур, текстура тощо.

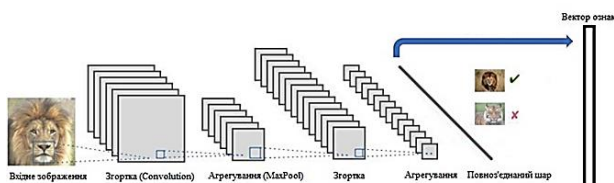


Рис. 3. Загальна архітектура мережі для перетворення зображення на вектор ознак

Отриманий від користувача пошуковий запит також обробляється нейронною мережею та будується вектор його ознак. Пошуковий запит може бути як у текстовій формі, так і у вигляді відповідного формату шуканого медіафайла – при пошуку зображення, порівнювати зі схожим зображенням; шукаючи аудіофайл, виконувати пошук за схожим відрізком звуку або схожою ритмічністю тощо.

Указаний підхід дозволяє зменшити витрати часу на аналіз файлу, але він потребує налаштування архітектури мережі для досягнення найбільшої точності пошуку. Оброблення великих векторів нейронною мережею потребує значних затрат ресурсів, тому зменшення розмірів векторів має значення для підвищення ефективності роботи системи при використанні цього методу.

Семантичний пошук. Після побудови анотації для мультимедіафайла, вона порівнюється з пошуковим запитом і визначається їхня відповідність. Існує два підходи до порівняння анотації та запиту – лексичний і семантичний. За використання лексичного підходу, відповідність оцінюють лише на основі лексичних збігів термінів із запиту в документі, тобто здійснюють пошук однакових слів. Такі класичні алгоритми для оцінювання важливості слова в контексті документа, як TF-IDF та Okapi BM25, є методами лексичного пошуку [9].

Але лексичний метод пошуку неефективний, якщо немає повного збігання слів у запиті та текстовому описі. Цю проблему розв'язує другий метод пошуку – семантичний. Семантичний пошук може віднаходити синоніми та споріднені слова до слів запиту, а також він стійкий до помилок у словах запиту.

Ефективне використання нейронних мереж для семантичного пошуку пов'язано з тим, що їхня глибока ієрархічна структура створює високі рівні абстракцій, аналізуючи необроблені дані, й автоматично визначає семантичні зв'язки. Загальну архітектуру нейромережі для виконання семантичного пошуку показано на рис. 3. Метою роботи даної мережі є представлення тексту запиту та тексту анотації файлу у вигляді таких векторів, скалярний добуток яких тим більший, чим більше релевантний документ до запиту – дані вектори називають векторами прихованих семантичних ознак. Перетворення вхідного тексту на вектори семантичних ознак відбувається за допомогою прихованих шарів. Для оцінювання релевантності документа до запиту, два отримані вектори семантичних ознак порівнюють один з одним, використовуючи деяку функцію подібності [9].

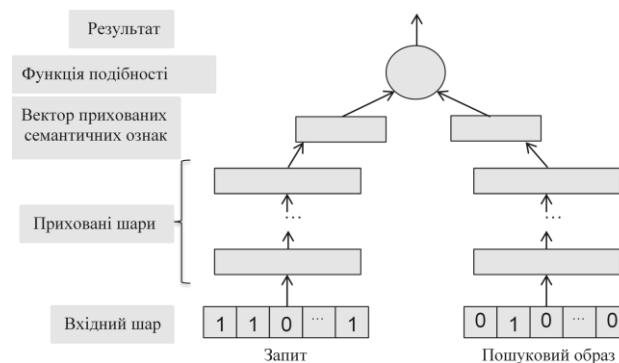


Рис. 4. Архітектура мережі для семантичного пошуку

Прикладом нейромережі для семантичного пошуку інформації є DSSM (Deep Semantic Similarity Model) [10]. Вхідні дані моделі DSSM – це багатовимірний вектор термінів у запиті або документі. Терміни розбиваються на триграми. Елементи вектора, що відповідають триграмам, присутнім у вхідному тексті та документі, набувають значення 1, в іншому випадку – 0. Вхідні вектори обробляються прихованими шарами, які перетворюють вхідний вектор на вектор прихованих семантичних ознак. Для обчислення оцінки релевантності між документами та запитом використовується функція косинусної подібності відповідних їм семантичних векторів понять [8]. Ціль навчання даної нейронної мережі – максимізувати вихідне значення функції подібності.

4. ВИСНОВОК

Великий обсяг мультимедіафайлів, викликав потребу в розробленні методів пошуку інформації серед таких файлів. Традиційні методи пошуку мультимедійної інформації, такі як, аналіз



гістограми кольорів зображення або частоти звуку, мають високу швидкість оброблення даних, але низьку точність через відсутність можливості виконання семантичного пошуку. Використання нейронних мереж дозволяє обробляти та шукати мультимедійну інформацію, а їхня здатність віднаходити семантичні зв'язки дає змогу виконувати семантичний пошук. Створення анотації мультимедіафайлів за допомогою нейронних мереж, підвищує точність і повноту пошуку, порівняно з традиційними методами. Але через неможливість створення нейромережі для класифікації всіх об'єктів реального світу, потрібно використовувати декілька нейромереж. Побудова вектора ознак файла за допомогою нейромереж також враховує його семантичне значення. Використовуючи вказаний підхід, можна будувати оптимальну мережу під конкретну задачу, що зменшить затрати ресурсів у разі побудови вектора, при цьому важливим є правильне налаштування нейронної мережі, адже це впливає на швидкість оброблення файла та точність пошуку.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

- [1] Sagarmay Deb. Multimedia Systems and Content-Based Image Retrieval. IGI Global, 2003, 406 p.
- [2] Gerald J. Kowalski, Mark T. Maybury. Information Storage and Retrieval Systems. Kluwer Academic Publishers, 2000, 333 p.
- [3] Athman Bouguettaya, Boualem Benatallah, Ahmed K. Elmagarmid. Interconnecting Heterogeneous Information Systems. Springer Science+Business Media, 1998, 228 p.
- [4] Yu-Jin Zhang. Semantic-Based Visual Information Retrieval. IRM Press, 2006, 385 p.
- [5] Zongmin Ma. Artificial Intelligence for Maximizing Content Based Image Retrieval. Information Science Reference, 2009, 429 p.
- [6] Alex Alemi. Improving Inception and Image Classification in TensorFlow. URL: <https://ai.googleblog.com/2016/08/improving-inception-and-image.html> (дата звернення: 7.02.2021).
- [7] Maha Mahmood, Wijdan Jaber Al-Kubaisy, Belal Al-Khateeb. Using Artificial Neural Network for Multimedia Information Retrieval. Journal of Southwest Jiaotong University. Vol 54, No 3 (2019).
- [8] Kostadinov S. Understanding GRU Networks. URL: <https://towardsdatascience.com/understanding-gru-networks-2ef37df6c9be> (дата звернення 7.02.2021).
- [9] Po-Sen Huang, Xiaodong He, Jianfeng Gao. Learning Deep Structured Semantic Models for Web Search using Clickthrough Data. At ACM International CIKM. October 2013.
- [10] Gia-Hung Nguyen, Lynda Tamine, Laure Soulier, Nathalie Souf. Toward a Deep Neural Approach for Knowledge-Based IR. URL: <https://arxiv.org/pdf/1606.07211.pdf> (дата звернення 7.02.2021).

Стаття надійшла до редколегії

10.10.2020



Searching for multimedia information based on neural networks

The article considers approaches to the use of neural networks in multimedia information retrieval. The development of methods for multimedia information retrieval is necessary due to the large amount of such information. Traditional methods of multimedia information retrieval have a high speed of data processing, but low accuracy due to the inability of semantic search. The use of neural networks allows for semantic search, which increases its accuracy and completeness. Approaches to the use of neural networks at the stages of indexing and retrieval of multimedia information are considered. With the help of a neural network, a multimedia file is analyzed and classified. The result of classifying a file is used to create its textual description - an annotation that is compared to the search query to determine relevance. There are many ready-made classification networks that can be used to speed up the process of creating a multimedia search system, but it is not possible to create a neural network to classify all real-world objects, so multiple neural networks should be used. Neural networks are also used to build feature vectors for a media file and a search query. Similarity functions, such as cosine of similarity, are applied to constructed vectors to determine the semantic similarity of a query and a media file. In this case, the search query can be both in text form and in the form of the appropriate format of the desired media file. This approach allows to build an optimal neural network for a specific task. Neural networks are used to compare the constructed annotation of a file and a query, which increases the accuracy and completeness of the search, compared to traditional methods, due to the ability of neural networks to take into account the semantic meaning of the text.

Keywords: neural networks; search for multimedia information; semantic search; information search systems; data annotation.



Богдан Бондаренко,
магістрант кафедри інтелектуальних технологій Київського національного університету імені Тараса Шевченка.

Bogdan Bondarenko,
Master's student of the Department of Intellectual Technologies of Taras Shevchenko National University of Kyiv.



Юрій Самохвалов,
доктор технічних наук, професор кафедри інтелектуальних технологій. Київський державний університет імені Тараса Шевченка.

Yuri Samokhvalov,
Doctor of Technical Sciences, Professor of the Department of Intellectual Technologies. Taras Shevchenko National University of Kyiv.