УДК 004.8
DOI: 10.31891/CSIT-2020-1-5

МАНЗЮК Е. А., СКРИПНИК Т. К., ГІРНИЙ М. Ю.
Хмельницький національний університет

ВИКОРИСТАННЯ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ РОЗПІЗНАВАННЯ СКЛАДОВИХ ЕЛЕМЕНТІВ ОБ'ЄКТІВ НА БАЗІ ЗОБРАЖЕННЯ

Розпізнавання зображень використовується для отримання, аналізу, розуміння і обробки зображень з реального світу, щоб перетворювати їх у цифрову інформацію. В цю область залучені інтелектуальний аналіз даних, машинне навчання, розпізнавання шаблонів, розширення бази знань.

Система розпізнання дозволяє зробити крок до систем розуміння продуктів харчування, таких як оцінка калорій та створення рецептів. Система розпізнання може бути застосована для вирішення більш широких проблем, таких як прогнозування зображення на встановлення відповідності складових елементів.

Ключові слова: класифікація, розпізнавання зображень, нейронні мережі

MANZIUK E., SKRYPNYK T., HIRNYI M.
Khmelnytskyi National University, Ukraine

DETERMINATION OF RECIPES CONSTITUENT ELEMENTS BASED ON IMAGE

Image recognition is used to retrieve, analyse, understand, and process images from the real world to convert them into digital information. In this area involved data mining, machine learning, pattern recognition, knowledge extension.

Developments in the image recognition area have resulted in computers and smartphones becoming capable of mimicking human eyesight. Improved cameras in modern devices can take pictures of very high quality, and with the help of new software, they receive the necessary information and on the basis of the received data is processed images.

However, food recognition challenges modern computer vision systems and needs to go beyond just an visible image. Compared to understanding the natural image, visual prediction of ingredients requires high-level solutions and previous knowledge. This creates additional problems, because food components have high variability between the class, when cooking, you have to convert components and the ingredients are often included in the cooked dish. The recognition system allows you to take a step toward understanding the food supply systems such as calorie score and create recipes. The recognition system can be used to address wider problems, such as the prediction of the image on the consistency of the folding elements.

Keywords: classification, image recognition, neural networks.

Вступ. Як і будь-яка інша сфера на самопочуття людини впливає і харчування. Щодня незліченні фотографії їжі публікуються від користувачів у соціальних мережах; від першого домашнього торта до страви високої кухні.

Успіхи в класифікації окремих кулінарних інгредієнтів рідкісні [1]. Проблема полягає в тому, що майже немає публічно відредагованих записів. У цій роботі розглядається проблема автоматизованого розпізнавання сфотографованої страви для приготування та подальший вихід відповідного рецепту. Відмінність складності обраної проблеми від інших контрольованих проблем класифікації полягає в тому, що в стравах є великі перекриття, так само як і велика внутрішньокласова схожість, оскільки страви різних категорій можуть виглядати дуже схоже лише за інформацією по зображенню але не схожих на смак.

Розробки в області розпізнавання зображень привели до того, що комп'ютери і смартфони стали здатні імітувати людський зір. Вдосконалені камери в сучасних пристроях можуть знімати фотографії дуже високої якості, а за допомогою нового програмного забезпечення з них отримується необхідна інформація і на базі отриманих даних відбувається обробка зображень.

Однак розпізнавання їжі кидає виклик сучасним системам комп'ютерного зору та потребує вийти за рамки просто видимого зображення. У порівнянні з розумінням природного зображення, візуальне передбачення інгредієнтів вимагає рішень високого рівня та попередніх знань. Це створює додаткові проблеми, оскільки харчові компоненти відрізняються високою мінливістю між класом, при готуванні виникають перетворення з компонентами, а інгредієнти часто включаються в приготовлену страву.

Аналіз досліджень та публікацій. Останні кілька років спостерігаються неабиякі вдосконалення таких завдань візуального розпізнавання, як класифікація природного зображення [2, 3], виявлення об'єктів [4, 5] та семантична сегментація [6, 7]. Однак, на відміну від розуміння природного зображення, розпізнавання їжі створює додаткові проблеми, оскільки їжа та її компоненти мають велику мінливість у внутрішньокласовій формі та мають сильні деформації, що виникають у процесі готування.

Використовуючи нейронні мережі [8], запропоновані рішення для одночасного вивчення розпізнавання інгредієнтів та категоризації харчових продуктів, використовуючи взаємні зв'язки між ними. Було вивчено смислові етикетки інгредієнтів, і потім особливості взаємозв'язків використовувались для пошуку рецептів.

Широке коло досліджень в яких було докладено значних зусиль для використання глибоких нейронних мереж для багатозначної класифікації шляхом розробки моделей [9–11], методи групової класифікації [12, 13] та вивчення функцій втрат [14], які, як показала практика досить добре підходять для

таких завдань. Застосування методів візуального визначення групованих об'єктів дозволяє відокремити групи класів [15–17]

Ряд досліджень вирішують досить складні завдання, такі як оцінка кількості калорій, наданих по зображенню їжі [18], оцінка кількості їжі [19], прогнозування списку наявних інгредієнтів [20] та пошук рецепту для даного зображення [21–23]. Крім того, [24] надає детальний аналіз харчових рецептів, враховуючи зображення, атрибути та інгредієнти рецептів. Завдання, пов'язані з харчовими продуктами, також розглядалися в літературі з обробки природних мов, де вивчення рецептів вивчалося в контексті генерування процесуального тексту [25] або контрольних списків інгредієнтів [26].

Система аналізу зображення. Відповідно до сучасного стану, було використано найбільший набір даних із понад 125000 рецептів (eightportions.com). Запропоновано поєднання розпізнавання об'єктів або розпізнавання страв з приготуванням за допомогою конволюційних нейронних мереж (CNN – Convolutional Neural Networks) та пошук найближчих сусідів (NNC – Next-Neighbor Classification) у записі зображень. Ця комбінація допомагає швидше знайти правильний рецепт, оскільки топ-5 категорій CNN порівнюються з категорією наступного сусіда (NNC) з ранговою кореляцією. Підходи, що базуються на кореляції рейтингу, такі як Кендалл Тау по суті, вимірюють ймовірність того, що два пункти будуть в одному порядку в двох ранжируваних списках.

Формула обчислення ***коефіцієнта рангової кореляції Кендалла*** може бути виражена як:

$$\tau = \frac{P(p) - P(q)}{N\dfrac{N-1}{2}}, \tag{1}$$

де $P(p)$ – кількість збігів;

$P(q)$ – кількість інверсій;

N – обсяг вибірки.

У ***спрощеному вигляді*** формулу коефіцієнта кореляції Кендалла можна записати як:

$$\tau = \frac{4P}{N(N-1)} - 1. \tag{2}$$

При наявності пов'язаних рангів формула змінюється з урахуванням поправки на пов'язані ранги:

$$\tau = \frac{P(p) - P(q)}{\sqrt{N\dfrac{N-1}{2} - K_X}\sqrt{N\dfrac{N-1}{2} - K_Y}}, \tag{3}$$

де $P(p)$ – кількість збігів;

$P(q)$ – кількість інверсій;

N – обсяг вибірки,

K_X – поправка на зв'язок рангів змінної X,

K_Y – поправка на зв'язок рангів змінної Y.

$$K_X = \sum_i |X|_i \left(|X|_i - 1\right);$$
$$K_Y = \sum_j |Y|_j \left(|Y|_j - 1\right), \tag{4}$$

де i – кількість груп зв'язків з X;

j – кількість груп зв'язків з Y.

Система розпізнавання має наступну структуру:

1. На кожен рецепт є певна кількість зображень. Для кожного з цих зображень функціональні вектори генеровані із заздалегідь підготовленої нейронною мережею, що навчається на 1000 категоріях в змаганні з розпізнавання зображень ILSVRC з мільйонами зображень. Функціональні вектори формують внутрішнє зображення в останньому повністю пов'язаному. Ці вектори функцій потім розмірно зменшуються за допомогою аналізу основного компонента (PCA – Principal Component Analysis) від матриці N × 4096 до матриці N × 512. Як результат, вибирається 5 найкращих зображень з найменшою евклідовою відстанню до вхідного зображення (приблизний найближчий сусід), тобто 5 найкращих оптичних, безпосередньо з інформації про зображення, подібних зображень до вхідного зображення.

2. Крім того, CNN навчається з кількістю категорій із зображеннями рецептів. Кількість категорій була визначена динамічно з допомогою теми і семантичного аналізу назв рецептів. У результаті отримуємо для кожної категорії ймовірність, згідно якої може належати вхідне зображення.

3. Найкращі категорії CNN (п. 2) порівнюються з категоріями з оптично схожих зображень з (п. 1) з кореляцією Кендалла.

Схема візуалізації методу виглядає приблизно так:

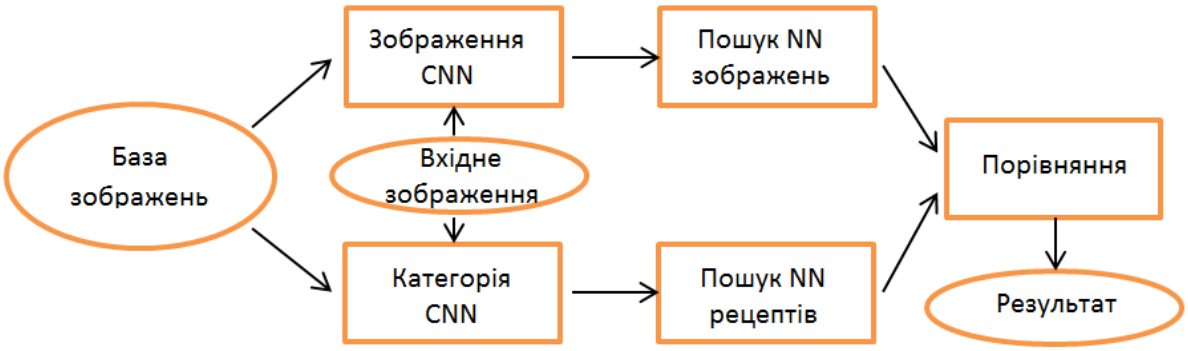

Рис. 1. Схема функціонування системи розпізнавання зображень

Загалом послідовний алгоритм за який функціонує система розпізнавання зображень можна представити в такому вигляді.

1 **Підготовка даних**
Очищення даних
Розширення даних
2 **Аналіз та візуалізація даних, розділені дані**
3 **Тематичне моделювання**
Латентне розподілення Діріхле (LDA).
Матрична факторизація.
4 **Добування особливостей**
k-найближчі сусіди.
Візуалізація t-SNE.
5 **Трансферне навчання: Навчання заздалегідь підготовленої CNN (конволюційна нейронна мережа)**
6 **Розгортання**

Для того, щоб взагалі можна було тренувати модель, потрібні достатньо даних (як засіб можна використовувати так зване нарощування даних та точне налаштування попередньо підготовлених моделей). Тільки завдяки цьому кількість узагальнення даних навчального набору може постійно збільшуватися до певної міри, і в тестовому наборі може бути досягнута висока точність.

Більше даних призводить до збільшення розмірів, але більше розмірів не обов'язково призводить до кращої моделі та її представлення. Відхилення шаблонів у наборі даних, які порушують навчання, можуть ненавмисно посилюватися більш великими розмірами, узагальнення та засвоєння запису даних погіршується для нейронної мережі, відношення сигнал-шум зменшується.

Наступним важливим кроком є вибір функцій для знецінення неважливих даних. Підготовка вихідних даних для нейронної мережі є звичною практикою. У першому запуску завантажується назва рецепту, середня заявка на рецепт, кількість оцінок, рівень складності, час підготовки та дата публікації. У другому проході потім список інгредієнтів, текст рецепта, всі зображення та кількість разів, коли рецепт був надрукований. За допомогою цих особливостей запис даних може бути описаний дуже добре і допомагає зрозуміти набір даних, що важливо для вибору алгоритмів. Такі дані, як назва рецепта, рейтинг, дата завантаження рецепта тощо, зберігаються у файлі. Для того, щоб отримати перше враження, ми зазвичай розробляємо теплову карту, щоб отримати першу інформацію, які можливі функції цікаві. Теплова карта дає нам зрозуміти, які значення співвідносяться з іншими значеннями.

Якщо видалити всі інгредієнти, які зустрічаються не один раз, залишаться унікальні. Для аналізу асоціацій інгредієнтів використовується алгоритм APRIORI. Це забезпечує частоту того, які інгредієнти в поєднанні з іншими інгредієнтами зустрічаються в загальній кількості, як часто.

Лідером інгредієнтів є сіль з 60-відсотковим представництвом у всіх рецептах. На третьому місці можна побачити перший кортеж, поєднання двох інгредієнтів, а саме перцю та солі з трохи більше 40 відсотків, вони, безумовно, найпоширеніша пара. Найпоширеніші трійні, чотиримісні та навіть четвірки/

Тематичне моделювання за категоріями. Мета цієї процедури - розділити всі назви рецептів на n-категорії. Для контрольованої проблеми класифікації необхідно надати нейронній мережі розмічені зображення. Лише за допомогою цих міток навчання стає можливим. Латентне розподілення Діріхле (LDA) – це вірогідна модель, яка передбачає, що кожне ім'я може бути призначене темі. По-перше, тіло імені повинно бути очищене, тобто слова зупинки видаляються, а слова зводяться до кореня. Чистий словник служить вводом. Як результат, є список ймовірностей того, наскільки певна модель, що вона відповідатиме темі.

Наступним кроком є обчислення tf-idf (термін обернена частота документа). Це значить лише важливість слова в назві рецепту , враховуючи його значення в усьому текстовому корпусі. Чотири найважливіші слова:

1. Салат (2673.14).
2. Спагеті (2368.45).
3. Торт (2045.12).
4. Торт (1430.58).

Результат можна візуалізувати за допомогою t-SNE. Важливо, щоб запис з кількома вимірами був зменшений до 2D, що дозволяє знайти координату для кожної назви рецепту.

Алгоритм NMF приймає як вхід tf-idf і одночасно виконує зменшення розмірів і кластеризацію. Це дає хороші результати, як зазначено нижче для перших 4 тем:

Тема № 0: спагеті карбонара олія.

Тема № 1: салат із дині цикорій редька селера.

Тема № 2: локшина китайська азіатський вок базилік.

Тема № 3: кекси чорничний фундук журавлина пікантна соковита шоколад

За допомогою мереж CNN інформація зображення спочатку узагальнюється для зменшення кількості параметрів. Припускаємо, що перші шари в CNN розпізнають шорсткі структури на малюнку. Чим далі переходимо до останнього шару, тим кращими стають функції розпізнання. Можемо скористатися цим і вибираємо заздалегідь підготовлені CNN, які були підготовлені мільйонами знімків, і видалити останні шари, щоб навчити їх за допомогою наших власних даних. Це економить нам мільйони параметрів і, таким чином, скорочує час на обчислення. Вибраний тут CNN – це VGG-16, який тренувався у змаганнях з класифікації 2014 року на 1000 категорій.

Якщо видалити останній шар, отримаємо витяжку функцій другого-останнього шару. Це утворює матрицю n x 4096, де n – кількість вхідних зображень.

Дозволяємо VGG-16 обчислювати вектор для кожного зображення, яке маємо. Цей вектор – відбиток малюнка: внутрішнє зображення, яке будує нейронна мережа.

Рис. 2. Вектор 4096 обчислений від зображення тістечка

Тепер все, що потрібно зробити – кожен новий заданий вхідний образ передати його через VGG-16, отримати вектор відбитків і обчислити найближчих сусідів з приблизним пошуком найближчого сусіда. Бібліотека, яку , було використано для цього, є FALCONN бібліотека з алгоритмами для пошуку найближчих сусідів. Алгоритми в програмі FALCONN засновані на локальному чутливому хешингу (LSH), який є популярним класом методів пошуку найближчого сусіда у великомірних просторах. Метою FALCONN є забезпечення дуже ефективних та добре перевірених реалізацій структур даних на основі LSH.

В даний час FALCONN підтримує два сімейства LSH для косинусної подібності: гіперплан LSH та поперечний багатогранник LSH. Обидві сім'ї хешів реалізовані за допомогою багатопробного LSH з метою мінімізації використання пам'яті. Крім того, FALCONN оптимізований як для щільних, так і для розріджених даних. Незважаючи на те, що створений для косинусної подібності, FALCONN часто можна використовувати для пошуку найближчого сусіда на евклідовій відстані або для максимального внутрішнього пошуку продукту.

Створення рецепту з зображення вимагає одночасного розуміння інгредієнтів, що складають страву, а також будь-якої обробки, яку вони пройшли, наприклад, нарізання або змішування з іншими інгредієнтами. Традиційно проблема «зображення до рецепта» була сформульована як завдання пошуку, де рецепт отримується з фіксованого набору даних на основі оцінки схожості зображення у вбудованому просторі. Продуктивність таких систем сильно залежить від розміру та різноманітності набору даних, а також від якості вивченого налаштування. Не дивно, що ці системи виходять з ладу, коли в статичному наборі даних не існує відповідного рецепту запиту зображення.

Альтернативою подолання обмежень набору даних систем пошуку є формулювання проблеми зображення-рецепта як умовної генерації. Замість отримання рецепту безпосередньо із зображення, система генерації рецептів отримала би перевагу від проміжного кроку: передбачення списку інгредієнтів.

Послідовність інструкцій потім формуватиметься як зображенням, так і відповідним списком інгредієнтів, де взаємодія між зображенням та інгредієнтами може дати додаткову інформацію про те, як останні оброблялися для отримання отриманого блюда.

Висновки. Система генерування зображень до рецептів приймає зображення їжі та видає рецепт, що містить назву, інгредієнти та інструкції з приготування. Метод починається з пошуку кодера зображення та декодера інгредієнтів, який прогнозує набір інгредієнтів, використовуючи візуальні особливості, витягнуті з вхідного зображення та спільного появи інгредієнтів. Потім тренуємо кодер інгредієнтів та декодер інструкцій, які генерують заголовок та інструкції, беручи до уваги візуальні особливості зображення та передбачувані інгредієнти та подаючи їх у сучасну модель генерації послідовностей. Таким чином система визначення інгредієнтів страви із зображення дозволяє з високою ймовірністю, за найкращими порівняннями отримати прийнятні результати автоматично визначення рецепту та вихідних складових елементів страви.

Література

1. Aizawa, K., & Ogawa, M. (2015). Foodlog: Multimedia tool for healthcare applications. IEEE MultiMedia, 22(2), 4-8.
2. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
3. He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In Proceedings of the IEEE international conference on computer vision (pp. 1026-1034).
4. Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems (pp. 91-99).
5. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).
6. Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3431-3440).
7. Jégou, S., Drozdzal, M., Vazquez, D., Romero, A., & Bengio, Y. (2017). The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 11-19).
8. Chen, J., & Ngo, C. W. (2016, October). Deep-based ingredient recognition for cooking recipe retrieval. In Proceedings of the 24th ACM international conference on Multimedia (pp. 32-41).
9. Wei, Y., Xia, W., Huang, J., Ni, B., Dong, J., Zhao, Y., & Yan, S. (2014). Cnn: Single-label to multi-label. arXiv preprint arXiv:1406.5726.
10. Nam, J., Mencía, E. L., Kim, H. J., & Fürnkranz, J. (2017). Maximizing subset accuracy with recurrent neural networks in multi-label classification. In Advances in neural information processing systems (pp. 5413-5423).
11. Wang, J., Yang, Y., Mao, J., Huang, Z., Huang, C., & Xu, W. (2016). Cnn-rnn: A unified framework for multi-label image classification. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2285-2294).
12. Barmak, O. V., Krak, Y. V., & Manziuk, E. A. (2018). Characteristics for choice of models in the ansables classification. PROBLEMS IN PROGRAMMING, (2-3), 171-179.
13. Manziuk, E. A., Barmak, A. V., Krak, Y. V., & Kasianiuk, V. S. (2018). Definition of information core for documents classification. Journal of Automation and Information Sciences, 50(4).
14. Gong, Y., Jia, Y., Leung, T., Toshev, A., & Ioffe, S. (2013). Deep convolutional ranking for multilabel image annotation. arXiv preprint arXiv:1312.4894.
15. Meyers, A., Johnston, N., Rathod, V., Korattikara, A., Gorban, A., Silberman, N., ... & Murphy, K. P. (2015). Im2Calories: towards an automated mobile vision food diary. In Proceedings of the IEEE International Conference on Computer Vision (pp. 1233-1241).
16. Krak, I., Barmak, O., & Manziuk, E. (2020). Using visual analytics to develop human and machine-centric models: A review of approaches and proposed information technology. Computational Intelligence.
17. Barmak, O., Manziuk, E., & Krak, I. (2019, September). Using piecewise hyper linear classification in multidimensional feature space for text content. In 2019 IEEE 14th International Conference on Computer Sciences and Information Technologies (CSIT) (Vol. 2, pp. 119-123). IEEE.
18. Barmak, A. V., Krak, Y. V., Manziuk, E. A., & Kasianiuk, V. S. (2019). Information technology of separating hyperplanes synthesis for linear classifiers. Journal of Automation and Information Sciences, 51(5).
19. Chen, M. Y., Yang, Y. H., Ho, C. J., Wang, S. H., Liu, S. M., Chang, E., ... & Ouhyoung, M. (2012). Automatic chinese food identification and quantity estimation. In SIGGRAPH Asia 2012 Technical Briefs (pp. 1-4).
20. Chen, J. J., Ngo, C. W., & Chua, T. S. (2017, October). Cross-modal recipe retrieval with rich food attributes. In Proceedings of the 25th ACM international conference on Multimedia (pp. 1771-1779).
21. Wang, X., Kumar, D., Thome, N., Cord, M., & Precioso, F. (2015, June). Recipe recognition with large multimodal food dataset. In 2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW) (pp. 1-6). IEEE.
22. Salvador, A., Hynes, N., Aytar, Y., Marin, J., Ofli, F., Weber, I., & Torralba, A. (2017). Learning cross-modal embeddings for cooking recipes and food images. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3020-3028).
23. Carvalho, M., Cadène, R., Picard, D., Soulier, L., Thome, N., & Cord, M. (2018, June). Cross-modal retrieval in the cooking context: Learning semantic text-image embeddings. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (pp. 35-44).
24. Min, W., Bao, B. K., Mei, S., Zhu, Y., Rui, Y., & Jiang, S. (2017). You are what you eat: Exploring rich recipe information for cross-region food analysis. IEEE Transactions on Multimedia, 20(4), 950-964.
25. Mori, S., Maeta, H., Sasada, T., Yoshino, K., Hashimoto, A., Funatomi, T., & Yamakata, Y. (2014, June). Flowgraph2text: Automatic sentence skeleton compilation for procedural text generation. In Proceedings of the 8th International Natural Language Generation Conference (INLG) (pp. 118-122).
26. Kiddon, C., Zettlemoyer, L., & Choi, Y. (2016). Globally coherent text generation with neural checklist models. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (pp. 329-339).

Надійшла / Paper received: 12.08.2020
Надрукована / Paper Printed : 02.09.2020