

ДИПЛОМНА РОБОТА МАГІСТРА

на тему Використання штучного інтелекту для оцінювання розпізнавання зображень

Галузь знань 12 – Інформаційні технології

Шифр і назва галузі знань

Спеціальність 122 – Комп'ютерні науки

Шифр і назва спеціальності

Виконав: студент 2 курсу, група КНМ-19-1



Підпис

М.Ю. Гірний

Ініціали, прізвище

Керівник: к.т.н., доцент кафедри КНІТ



Підпис

Е.А. Манзюк

Ініціали, прізвище

Нормоконтроль: к.т.н., доцент кафедри КНІТ



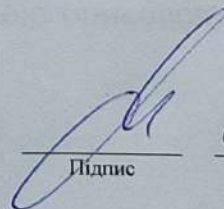
Підпис

Р.О. Багрій

Ініціали, прізвище

До захисту допускаю:

Зав. кафедри КНІТ, д.т.н., професор



Підпис

О. В. Бармак

Ініціали, прізвище

7 12 2020 р.

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

Факультет програмування та комп'ютерних і телекомунікаційних систем

Кафедра комп'ютерних наук та інформаційних технологій

Освітній ступінь магістр

Галузь знань 12 – Інформаційні технології

Спеціальність 122 – Комп'ютерні науки

ЗАТВЕРДЖУЮ

Завідувач кафедри комп'ютерних наук та інформаційних технологій


(підпис)

д.т.н., професор О.В. Бармак

« 7 » 9 2020 року

**ЗАВДАННЯ
НА ДИПЛОМНУ РОБОТУ МАГІСТРА**

1. Тема дипломної роботи магістра: «Використання штучного інтелекту для оцінювання розпізнавання зображень»

2. Завдання видано студенту Гірний Микола Юрійович
(прізвище, ім'я, по батькові)

3. Керівник роботи к.т.н., доцент Манзюк Едуард Андрійович
(прізвище, ім'я, по батькові)

4. Затверджені наказом університету від « 9 » 9 2020 р. № 22

5. Зміст пояснювальної записки (перелік задач) та вихідні дані:

Мета роботи – розробка методу отримання інформації для систем автоматичного формування рецептів базуючись на зображенні блюда. Об'єктом дослідження є методи отримання інформації із зображень та описової текстової інформації. Предметом дослідження є зображення блюд, а також описи рецептів та відповідні інгредієнти, що є додатковою описовою інформацією до зображень.

Реферат

Дипломна робота магістра присвячена використанню штучного інтелекту для оцінювання розпізнавання зображень.

Актуальність теми. В магістерській роботі розроблений і реалізований метод отримання інформації для розпізнавання рецептів та інгредієнтів за зображенням страв та їх рецептурою. Розроблювальна система пропонує користувачам страви на основі виявлених схожих, які виражені в текстовій формі та у вигляді зображень. Така система дозволяє за зображенням згенерувати рецепт та скласти список інгредієнтів базуючись на подібності зображення та схожості найближчих рецептів.

Метою дослідження є розробка методу отримання інформації для систем автоматичного формування рецептів базуючись на зображенні страви.

Для досягнення зазначеної мети поставлені наступні **задачі**:

- показати, що використання візуальної інформації та додаткової текстової інформації дозволяє створити систему розпізнавання рецептів а, відтак і інгредієнтів, базуючись на зображенні страви;
- провести дослідження ознак і класифікаторів, застосовуваних в машинному навчанні в задачах розпізнавання складових елементів та інтерпретації зображень;
- провести порівняння застосовності методів машинного навчання, до задачі отримання інформації про зображення базуючись на подібності зображення та подібності описової текстової інформації, в даному випадку рецептів.

При цьому передбачається розв'язок таких **підзадач**, як

- попередня обробка колекції рецептів та зображення страв;
- побудова словників рецептів;
- побудова шаблонів рецептурних форм та зображень та розміщення їх за групами;

- побудова кластерів подібних страв за зображенням та описовими формами - інгредієнтами;
- розробка ймовірнісних моделей групування зображень;
- виділення тих параметрів ресторанів, для отримувannya яких може бути застосоване машинне навчання (на основі розподілу значень параметрів в розміченому підкорпусі);
- вибір моделей, виділення ознак і застосування методів машинного навчання для отримувannya інформації про відповідні параметри;
- тестування методів на основі правил із використанням машинного навчання та вибір оптимального методу для кожного з параметрів;
- програмна реалізація модуля отримувannya інформації та думок з відгуків про заклади громадського харчування для рекомендаційної системи.

Об'єктом дослідження є методи отримувannya інформації із зображень та описової текстової інформації.

Предметом дослідження є зображення страв, а також описи рецептів та відповідні інгредієнти, що є додатковою описовою інформацією до зображень.

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.

Достовірність результатів забезпечується проведенням всебічного оцінювання та порівняння ефективності різних методів, а також на візуальному порівнянні та якості рекомендованого рішення.

Система генерувannya зображень по рецептам 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 спільної появи інгредієнтів. Потім тренуємо кодер інгредієнтів та декодер інструкцій, які генерують заголовок та інструкції, беручи до уваги візуальні особливості зображення та передбачувані інгредієнти подаючи їх у модель генерації послідовностей.

Практична значимість дослідження полягає в тому, що описані методи та отримані результати можуть бути використані при розробці систем автоматичного генерування рецептів на базі зображень.

В результаті проведеного дослідження показано, що за допомогою даних, отриманих на основі аналізу зображень та рецептурного опису, може бути підвищена ефективність роботи системи у розпізнанні зображень базуючись також на додатковому описі тобто текстовою інформацією до зображення.

Що стосується експериментів з машинним навчанням, в роботі показано експериментальні дослідження запропонованого методу навчання базуючись на зображенні та текстовому описі.

Апробація дипломної роботи.

Основні положення і результати роботи опубліковані в збірнику наукових праць – Манзюк Е.А. Використання штучного інтелекту для розпізнавання складових елементів об'єктів на базі зображення / Е.А. Манзюк, Т.К. Скрипник, М.Ю. Гірний // Прийнята до друку Вісник Хмельницького національного університету.

Структура та обсяг роботи. Дипломна робота магістра складається з завдання, реферату, змісту, вступу, 4 розділів, висновків, переліку посилань із 26 найменувань та додатків. Загальний обсяг дипломної роботи магістра становить 72 сторінок, з них 72 сторінка основного тексту та 1 сторінки додатків. в роботі наведено 26 рисунків.

Ключові слова: нейроні мережі, класифікація зображень, машинне навчання.

Зміст

Вступ.....	8
Розділ 1	12
Аналіз сучасного стану розпізнавання зображень	12
1.1 Опис предметної області	12
1.2 Визначення завдань розпізнавання.....	13
1.3 Системи розпізнавання на базі нейронних мереж	17
1.4 Нейронні мережі Хопфілда.....	19
1.5 Постановка задачі.....	21
Висновок до розділу 1.....	22
Розділ 2	24
Система аналізу зображення	24
2.1 Функціонування системи розпізнавання зображень	24
2.2 Структура системи розпізнавання.....	25
2.3 Тренування системи розпізнавання	27
2.4 Визначення чутливості системи до якості зображення	31
Висновки до розділу 2	34
Розділ 3	35
Розробка інформаційної технології розпізнавання зображень	35
3.1 Формування згорткової нейронної мережі	35
3.1 Функції активації.....	36
3.2 Функція втрат	36
3.4 Архітектура згорткової нейронної мережі.....	38
3.5 Зв'язний шар.....	38
3.6 Згортковий шар	39
3.7 Моделі використання згорткової нейронної мережі	42
Висновки до розділу 3	47
Розділ 4	48

Дослідження роботи системи розпізнавання.....	48
4.1 Трансформування даних.....	48
4.2 Збір і підготовка даних.....	51
4.3 Аналіз даних і візуалізація.....	55
4.4 Моделювання тем.....	58
4.5 Латентний розміщення Діріхте.....	59
4.6 Не негативна матрична факторизація.....	61
Висновки до розділу 4.....	66
Загальні висновки.....	67
Перелік посилань.....	69
Додатки	

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.

Актуальність теми. В магістерській роботі розроблений і реалізований метод отримання інформації для розпізнавання рецептів та інгредієнтів за

зображенням страв та їх рецептурою. Розроблювальна система пропонує користувачам страви на основі виявлених схожих, які виражені в текстовій формі та у вигляді зображень. Така система дозволяє за зображенням згенерувати рецепт та скласти список інгредієнтів базуючись на подібності зображення та схожості найближчих рецептів.

Метою дослідження є розробка методу отримання інформації для систем автоматичного формування рецептів базуючись на зображенні страви.

Для досягнення зазначеної мети поставлені наступні **задачі**:

- показати, що використання візуальної інформації та додаткової текстової інформації дозволяє створити систему розпізнавання рецептів а, відтак і інгредієнтів, базуючись на зображенні страви;
- провести дослідження ознак і класифікаторів, застосовуваних в машинному навчанні в задачах розпізнавання складових елементів та інтерпретації зображень;
- провести порівняння застосовності методів машинного навчання, до задачі отримання інформації про зображення базуючись на подібності зображення та подібності описової текстової інформації, в даному випадку рецептів.

При цьому передбачається розв'язок таких **підзадач**, як

- попередня обробка колекції рецептів та зображення страв;
- побудова словників рецептів;
- побудова шаблонів рецептурних форм та зображень та розміщення їх за групами;
- побудова кластерів подібних страв за зображенням та описовими формами - інгредієнтами;
- розробка ймовірнісних моделей групування зображень;

- виділення тих параметрів ресторанів, для отримання яких може бути застосоване машинне навчання (на основі розподілу значень параметрів в розміченому підкорпусі);
- вибір моделей, виділення ознак і застосування методів машинного навчання для отримання інформації про відповідні параметри;
- тестування методів на основі правил із використанням машинного навчання та вибір оптимального методу для кожного з параметрів;
- програмна реалізація модуля отримання інформації та думок з відгуків про заклади громадського харчування для рекомендаційної системи.

Об'єктом дослідження є методи отримання інформації із зображень та описової текстової інформації.

Предметом дослідження є зображення страв, а також описи рецептів та відповідні інгредієнти, що є додатковою описовою інформацією до зображень.

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.

Достовірність результатів забезпечується проведенням всебічного оцінювання та порівняння ефективності різних методів, а також на візуальному порівнянні та якості рекомендованого рішення.

Система генерування зображень по рецептам 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 спільної появи інгредієнтів. Потім тренуємо кодер інгредієнтів та декодер інструкцій, які генерують заголовок та інструкції, беручи до уваги візуальні особливості зображення та передбачувані інгредієнти подаючи їх у модель генерації послідовностей.

Практична значимість дослідження полягає в тому, що описані методи та отримані результати можуть бути використані при розробці систем автоматичного генерування рецептів на базі зображень.

В результаті проведеного дослідження показано, що за допомогою даних, отриманих на основі аналізу зображень та рецептурного опису, може бути підвищена ефективність роботи системи у розпізнанні зображень базуючись також на додатковому описі тобто текстовою інформацією до зображення.

Що стосується експериментів з машинним навчанням, в роботі показано експериментальні дослідження запропонованого методу навчання базуючись на зображенні та текстовому описі.

Розділ 1

Аналіз сучасного стану розпізнавання зображень

1.1 Опис предметної області

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 вивчалось в контексті генерування процесуального тексту [25] або контрольних списків інгредієнтів [26].

1.2 Визначення завдань розпізнання

Завдання розпізнавання (точніше, класифікування) об'єкта полягає в наступному. Існує якийсь спосіб кодування об'єктів (таких як рукописні літери), що належать до попередньо відомого кінця світу класу C_0C_1, \dots, C_q , і деяких об'єктів, кожен з яких відомий, до якого класу він належить. Вам потрібно створити алгоритм, який для будь-якого вхідного об'єкта, який не обов'язково належить до навчального набору, вирішує, до якого класу належить, і робить це досить добре. Якість розпізнавання оцінюється як ймовірність (тобто частоти) помилки класифікації на іншому кінцевому наборі об'єктів з попередньо відомими відповідями (тестовим набором).

Типова система розпізнавання складається з трьох частин: показу знаків, розпізнавання і прийняття рішення.

Показ знаків перетворює входи в рівномірний, компактний і зручний вигляд, з втратою переважної більшості інформації, що міститься в об'єкті, що мало впливає на класифікацію. Зручно виходить представлення об'єкта за формою стандарту Euclidean Space R_d , що відноситься до якогось фіксованого компактного значення (куб, м'яч, сфера, ...). Розмір d повинен бути досить великим для успішного (з точки зору якості) визначення і досить невеликим для успішного (в сенсі швидкості) визначення - насправді це близько декількох десятків. Спосіб отримання залежить від характеру і оригінального кодування об'єктів і знаходиться вручну. Наприклад, траєкторія миші або пера, спочатку закодована послідовністю довільної довжини (близько сотні), що складається з пари точкових координат, зручно кодувати послідовністю фіксованої довжини пар коефіцієнтів, що приближують траєкторію поліномів невеликого ступеня

(близько десятка), і навіть вільні члени можуть бути відкинуті як такі, що не впливають на класифікацію. Відомі бажані значення F в точках простору, що відповідають навчальному набору, тому залишається тільки побудувати в певному сенсі наближення відображення. Якість наближення буде перевірена не на всій ділянці визначення, а тільки на тестовому наборі. Інтерпретація обчислюваних ймовірностей здійснюється за іншою від розпізнавання процедурою прийняття рішень, яка будується вручну і не залежить ні від характеру вхідних об'єктів, ні від простору знаків, ні від навчальних даних. Це залежить тільки від того, для чого розрахована ця система розпізнавання. Наприклад, якщо визначається як невідомий клас, просто видається кількість найбільш вірогідних класів. Якщо об'єкт визначається як невідомий клас, то видає кількість найбільш вірогідних класів, якщо його ймовірність значно перевищує ймовірність інших класів, а відповіді «не знають» відсутні. Якщо використовується для генерації гіпотез, виробляється кількість декількох (наприклад, п'ять) найбільш імовірних класів і їх ймовірність.

У минулому столітті на уроках чистого письма першокласників спочатку змушували малювати палиці, гачки і кола, а вже потім вчили створювати з них листи. У реальному житті люди також пишуть літери як послідовності дрібних типових елементів, але не такі однорідні, як у записах. Наприклад, дві вертикальні палички, перекреслюючись горизонтально, ймовірно, є буквою "Н". І якщо вертикальні палички не зовсім палиці або не зовсім вертикальні, це все одно "Н". Але якщо вони наближаються зверху, це більше схоже на "А", а якщо горизонтальна палиця висока, то на "Р".

При розпізнаванні машини зручніше виділяти не тільки і не стільки палички і гачки, скільки способи стикування їх між собою: спрямовані в різні боки кліщі, петлі, Т-образні гілки і т. д. Припустимо, що дана чернетка роботи виконана, а вивчений фрагмент рукописного тексту закодований у вигляді послідовності таких характерних елементів, вправо впорядкованих як щось

зверху вниз. Питання в тому: наскільки схожа ця послідовність з буквою "А". Те ж питання для всіх інших впізнаваних об'єктів.

Спробуємо імітувати побудову однієї літери у вигляді послідовності елементів. По-перше, лист може мати кілька істотно різних правописів (наприклад, трикутний "А", трикутний "А" з завитками, "А" круглий, ...), і вони повинні бути змодельовані окремо. По-друге, навіть одне написання однієї літери, в залежності від почерку, генерує різні послідовності характерних елементів. По-третє, за винятком правильних творів неправильно (рука змістилась, крапка, ручка погано пише, ...). Всі можливості не передбачені, тому зможемо лише кілька найпоширеніших варіантів і що виявляється невеликими зміщеннями.

Виписуючи послідовність Q_{i1}, \dots, Q_{iL} з характерних елементів L , отримаємо детерміновану кінцеву машину з станами $L+1$ c_1, \dots, c_{L+1} , яка в кожному i -у, $i \in L$, стану c_L генерує елемент Q_{iL} і переходить до наступного стану. А щоб генерувати можливу послідовність «малих збурень», перетворимо машину в стохастичну: в i -у вона може генерувати будь-який елемент Q_j з імовірністю b_{ij} і з імовірністю a_{ij} перейти до стану, близького до $i+1$. Як правило $a_{ij} > 0$, тільки коли $i < j < i+2$, тобто на додаток до переходу до наступного стану, дозволяється отримати і затримати поточний стан. Важливо, щоб $a_{ij} = 0$ на $i > j$. Ці моделі називаються моделями LR (зліва-справа, зліва направо). Ймовірності b_i , a_i та інші позитивні ймовірності близькі до 0, хоча це не обов'язково. Зауважимо, що кожен стан детермінованої машини мав точне значення («видано множині елементів такої і такої послідовності»), і стохастична машина це точне значення.

Машини для декількох варіантів написання символів об'єднуються в одну машину, моделюючи символ: до їх неузгодженого поєднання додається загальний початковий стан C_0 , який не генерує жодного характерного елемента і від якого можливі переходи до початкового і других станів всіх варіантів, а їх скінченні стани склеюються між собою.

Ця модель символів дозволяє відповісти на питання, що якась послідовність характерних елементів схожа на цей символ: стільки ж, скільки ймовірність знаходження його моделі. Ці ймовірності можна обчислити для всіх символів і проаналізувати, щоб вирішити, який символ найбільш схожий на послідовність, тобто класифікувати її.

Розпізнавання послідовності об'єктів (наприклад, слів) суттєво відрізняється від розпізнавання окремих об'єктів (наприклад, букв) тим, що існує так багато різних довгих послідовностей, що навчання розпізнавати кожну послідовність особисто займає абсолютно неприйнятний час. Ще більш очевидним прикладом є розпізнавання молекул ДНК, які є дуже довгими словами в чотиримірному алфавіті. Таким чином, для розпізнавання послідовностей використовується наведена нижче схема.

- Послідовності більш-менш успішно сегментовані на окремі об'єкти або частини об'єктів. Алгоритми сегментації значно залежать від характеру об'єктів.

- Кожен об'єкт (або набір суміжних частин) більш-менш успішно розпізнається, наприклад, одним із методів, описаних у розділі.

- З відповідей розпізнавання об'єктів склеюється результат розпізнавання послідовності в цілому. Алгоритми склеювання можуть бути специфічними для цих об'єктів (наприклад, можна використовувати словник для розпізнавання слів відомою мовою, а за допомогою словника можна розпізнавати слова невідомою мовою, або вони можуть бути універсальними.

Тривіальний алгоритм розпізнавання - сегмент, розпізнавання кожного твору і як загальна відповідь, щоб дати послідовність найбільш імовірних відповідей - працює погано. Він не виправляє помилки сегментації і є адекватним лише визнанню випадкової послідовності незалежних добре розділених об'єктів. Але у всіх завданнях розпізнавання послідовності не випадкові, а в разі рукописного тексту - теж погано розділені.

1.3 Системи розпізнання на базі нейронних мереж

Теорія нейронних мереж виникла в 1940-х і 1960-х в результаті спільних спроб фізіологів і кібернетики зрозуміти і моделювати функцію мозку. Вийшла наступна модель: мозок складається з дуже великої кількості (близько 1011 далі млрд.) клітин (нейронів), кожна з яких має довгий хвіст (аксон) і велику кількість (близько 104) гілок (дендрити) щодо аксонів інших нейронів і / або вхідних рецепторів. Інформація (електрохімічний потенціал) може передаватися через ці сенсорні зони (синапси).

Кожен нейрон - це простий комп'ютер: потенціал нейрона (і його аксон, який грає роль виводу) - це функція потенціалу синапсів його дендрити (входів), і функція дуже специфічного виду. Зокрема, кожен нейрон має два стабільних стани (збуджені і незбуджені) і їх відповідні потенційні значення, однакові для всіх нейронів. Кожен нейрон обчислює лінійну комбінацію потенціалів вхідного синапсу, порівнює його з порогом і переходить у збуджений (незбуджений) стан, якщо лінійна комбінація більша (відповідно, менша) порогу.

Коллективно мозок обчислює якусь векторну функцію: залежність потенціалів нейронів (досить враховувати не всі нейрони, а тільки пов'язані з виконавцями) від потенціалів вхідних рецепторів. І вся не тривіальність функції мозку полягає в тому, що пороги (по одному на нейрон, в цілому близько 1011) і співвідношення лінійних комбінацій (по одному на дендрити, в цілому близько 1015), взагалі кажучи, різні і можуть змінюватися з плином часу. Ця зміна шансів називається навчанням. Пряме поширення нейронної мережі - це орієнтований ациклічний граф з багатьма V -верхами і ребрами E , верхівки яких розбиті на шари наступним чином:

- нульовий шар складається з вершин походження (вхідних рецепторів)

$V_{0,1}, \dots, V_{0,d}$;

- ребра (синапси), що походять у верхівки (нейрони) $(k+1)$ шару, виходять з верхівок (рецепторів або нейронів) k -шару;

- всі стоки (вихідні нейрони) $v_{L,1}, \dots, v_{L,q}$ відносяться до одного шару L -го.

Існує класифікація нейронних мереж.

1. Багат шарові нейронні мережі.

Архітектура багат шарової нейронної мережі (MNS) складається з послідовно з'єднаних шарів, де нейрон кожного шару з'єднаний його входами з усіма нейронами попереднього шару, а виходи - наступними. Система з двома вирішальними шарами може приблизно наближенано будь-якої багатовизначної функції з будь-якою точністю. Система з одним вирішальним шаром здатна утворювати лінійні розділові поверхні, що значно звужує спектр завдань, які вони вирішують, зокрема така мережа не зможе вирішити проблему типу «винятковий або». НМ з нелінійними функціями активації і двома вирішальними шарами дозволяє утворювати будь-які ділянки рішень в просторі прийняття рішень, а з трьома вирішальними шарами - ділянками будь-якої складності, в тому числі незадіяними.

При цьому не втрачає узагальненої здатності. MNS навчаються за допомогою алгоритму зворотного поширення помилки, яка є методом градієнтного спуску у ваговому просторі з метою мінімізації загальної похибки мережі. При цьому помилки (точніше величина корекції ваги) поширюються в зворотному напрямку від входів до виходів, через ваги, що з'єднують нейрони.

2. Нейронні мережі високого порядку.

Нейронні мережі високого порядку (NSPs) відрізняються від MNS тим, що вони мають лише один шар, але входи нейронів такі ж, як і показники високого порядку, які є продуктом двох або більше компонентів вхідного вектора. Такі мережі також можуть утворювати складні розділяючі поверхні. Особливість такої мережі в тому, що для навчання якогось класу досить представити свій образ без варіацій масштабу і поворотів - після тренування

мережа буде визначати відомі класи, незмінно масштабувати і перевертати. Така мережа не повністю пов'язана, швидко тренується і працює. Відбулося значне поліпшення точності класифікації такою мережею вбудованих і масштабних зображень в порівнянні з MNF.

1.4 Нейронні мережі Хопфілда

NS Hopfield (NSS) одношаровий і повністю підключений елемент, його виходи пов'язані з входами. На відміну від MNS, NSS - це релаксація - тобто встановлюється в початковому стані, функціонує до тих пір, поки не досягне стабільного стану, який буде його вихідним значенням. NSS використовується як асоціативна пам'ять і для оптимізації завдань. У першому випадку ННС тренується без вчителя (наприклад, за правилом Гебба), у другому випадку вага між нейронами спочатку закодована проблемою, яку потрібно вирішити. NSS синхронно, коли всі нейрони та асинхронні враховуються одночасно, коли перераховується випадково вибраний нейрон. Методи Ляпунова використовуються для вивчення динаміки мереж.

Показано, що асинхронний шар (NX) завжди сходиться до стабільних точок, а елементи синхронного NX є стабільними стаціонарними точками і обмежують цикли довжиною два шари. Таким чином, система від початкового стану сходиться до найближчого локального мінімуму мережевої енергії, стан нейронів в якому буде відновлено таким чином для завдань розпізнавання, а рішення - для оптимізаційних завдань. Стохастика модифікації NSS використовуються для пошуку глобального мінімуму для оптимізації завдань.

1.4.1 Самоорганізаційні нейронні мережі Кохонена

Самоорганізовані нейронні мережі Kohonen (SNC) забезпечують топологічне впорядкування вхідного простору зображень. Вони дозволяють

топологічно безперервно відобразити вхідний n -вимірний простір у вихідному m -вимірі.

Нейронна мережа з функцією радіальної основи (NSRBF) є подальшим розвитком НС Кохонен, в якій після конкурентного шару додали ще один шар, навчений методом зворотного розподілу. На відміну від НС Kohonen в NSRBF, виходи нейронів конкурентного шару є значеннями функції Гауса з нормальним законом розподілення, і потрібним обнулення нейронів. Ширина радіально-базової функції характеризує відстань між центром скупчення, яка утворюється кожним нейронним елементом і його найближчими сусідами.

1.4.2 Когнітрон

Когнітрон за своєю архітектурою схожий на структуру зорової кори, має ієрархічну багат шарову організацію, в якій нейрони між шарами з'єднуються тільки локально. Він вивчає конкурентне навчання (без вчителя). Кожен шар мозку усвідомлює різні рівні узагальнення; Вхідний шар чутливий до простих зображень, таких як лінії, і їх орієнтація в певних областях візуальної області, в той час як реакція інших шарів є більш складною, абстрактною і незалежною від положення зображення. Подібні функції реалізуються в когнітроні шляхом моделювання організації зорової кори.

Неокогнітрон є подальшим розвитком ідеї когнітрону і більш точно відображає структуру візуальної системи, дозволяє розпізнавати зображення незалежно від їх перетворень, обертань, перекосів і змін масштабу. Неогнітрон може як самостійно вчитися, так і вчитися у вчителя. Неокогнітрон отримує двовимірні зображення на вході, схожі на ті, що на сітковій оболонці ока, і обробляє їх наступними шарами так само, як було знайдено в зоровій корі людини. Звичайно, в неогнітроні немає нічого, що обмежує його використання тільки для обробки візуальних даних, воно досить універсальне і може широко використовуватися як узагальнена система розпізнавання шаблонів.

Вузли, що реагують на такі елементи, як лінії і кути певної орієнтації, були знайдені в зоровій корі. На більш високих рівнях вузли реагують на більш складні та абстрактні зображення, такі як кола, трикутники та прямокутники. На ще більш високих рівнях ступінь абстракції збільшується до тих пір, поки не будуть виявлені вузли, які реагують на обличчя і складні форми. Загалом вузли на більш високих рівнях отримують вхідні дані з групи низькорівневих вузлів і тому реагують на ширшу область поля зору. Вузли вищого рівня менш залежні від позицій і більш стійкі до спотворень.

Неокогнітрон є потужним засобом розпізнавання зображень, але вимагає високих обчислювальних витрат, які в даний час недосяжні.

У штучних нейронних мережах, як і в головному мозку, всі розрахунки відбуваються паралельно, і, таким чином, дуже швидко. Насправді нейронні мережі моделюються на звичайних послідовних комп'ютерах і працюють досить повільно, тому кількість верхівок і ребер мережі досягає збереження. У 1980-х на хвилі підвищеного інтересу до паралельних обчислень були створені досить активні апаратні впровадження нейронних мереж.

1.5 Постановка задачі

Метою дослідження є розробка методу отримання інформації для систем автоматичного формування рецептів базуючись на зображенні страви.

Для досягнення зазначеної мети поставлені наступні задачі:

- показати, що використання візуальної інформації та додаткової текстової інформації дозволяє створити систему розпізнавання рецептів а, відтак і інгредієнтів, базуючись на зображенні страви;
- провести дослідження ознак і класифікаторів, застосовуваних в машинному навчанні в задачах розпізнавання складових елементів та інтерпретації зображень;

- провести порівняння застосовності методів машинного навчання, до задачі отримування інформації про зображення базуючись на подібності зображення та подібності описової текстової інформації, в даному випадку рецептів.

При цьому передбачається розв'язок таких підзадач, як

- попередня обробка колекції рецептів та зображення страв;
- побудова словників рецептів;
- побудова шаблонів рецептурних форм та зображень та розміщення їх за групами;
- побудова кластерів подібних страв за зображенням та описовими формами - інгредієнтами;
- розробка ймовірнісних моделей групування зображень;
- виділення тих параметрів ресторанів, для отримування яких може бути застосоване машинне навчання (на основі розподілу значень параметрів в розміченому підкорпусі);
- вибір моделей, виділення ознак і застосування методів машинного навчання для отримування інформації про відповідні параметри;
- тестування методів на основі правил із використанням машинного навчання та вибір оптимального методу для кожного з параметрів;
- програмна реалізація модуля отримування інформації та думок з відгуків про заклади громадського харчування для рекомендаційної системи.

Висновок до розділу 1

Таким чином нейронні мережі не запрограмовані в звичному розумінні цього слова, вони навчені. Навчання є однією з головних переваг нейронних

мереж над традиційними алгоритмами. Технічно навчання полягає в тому, щоб знайти співвідношення зв'язків між нейронами. Під час навчання нейронна мережа здатна виявляти складний зв'язок між входом і виходом, а також виконувати узагальнення. Це означає, що в разі успіху мережа зможе повернути правильний результат на основі даних, яких не було в навчальному зразку.

Вихідні дані перетворюються на тип, в якому їх можна подавати до мережеских входів. Кожен запис у файлі даних називається навчальною парою або вектором навчання. Вектор навчання містить одне значення для кожного входу в мережу і, залежно від типу навчання (з вчителем або без нього), по одному значенню для кожного виходу з мережі. Навчання мережі на «сирому» наборі, як правило, не дає якісних результатів. Існує ряд способів поліпшити «сприйняття» мережі.

Нормування здійснюється, коли різні входи користуються даними різних вимірів. Наприклад, перший запис мережі подається значення від нуля до одного, а другий - від ста до тисячі. За відсутності нормування значення на другому вході завжди сягатимуть значно більшого впливу на вихід мережі, ніж значення на першому вході. При нормалізації об'єднуються розміри всіх вхідних і вихідних даних.

Розділ 2

Система аналізу зображення

2.1 Функціонування системи розпізнавання зображень

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, такого як Кендалл Тау, вимірюють ймовірність того, що два пункти будуть в одному порядку в двох ранжованих списках.

Формула обчислення коефіцієнта рангової кореляції Кендалла може бути виражена як:

$$\tau = \frac{P(p) - P(q)}{N \frac{N-1}{2}}, \quad (2.1)$$

де $P(p)$ - кількість збігів;

$P(q)$ - кількість інверсій;

N - обсяг вибірки.

У спрощеному вигляді формулу коефіцієнта кореляції Кендалла можна записати як:

$$\tau = \frac{4P}{N(N-1)} - 1 \quad (2.2)$$

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

$$\tau = \frac{P(p) - P(q)}{\sqrt{N \frac{N-1}{2} - K_X} \sqrt{N \frac{N-1}{2} - K_Y}} \quad (2.3)$$

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 .

$$\begin{aligned} K_X &== \sum_i |X|_i (|X|_i - 1), \\ K_Y &== \sum_j |Y|_j (|Y|_j - 1), \end{aligned} \quad (2.4)$$

де i - кількість груп зв'язків з X ;

j - кількість груп зв'язків з Y .

2.2 Структура системи розпізнавання

Система розпізнавання має наступну структуру. На кожен рецепт є певна кількість зображень. Для кожного з цих зображень функціональні вектори згенеровані із задалегідь підготовленої нейронною мережею, що навчається на 1000 категоріях в змаганні з розпізнаванням зображень з мільйонами зображень. Функціональні вектори формують внутрішнє зображення в останньому повністю пов'язаному векторі. Ці вектори функцій потім розмірно зменшуються за допомогою аналізу основного компонента (PCA - Principal Component Analysis) від матриці $N \times 4096$ до матриці $N \times 512$. 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.

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

Найкращі категорії CNN порівнюються з категоріями з оптично схожих зображень з кореляцією Кендалла.

Схема візуалізації методу виглядає приблизно так:

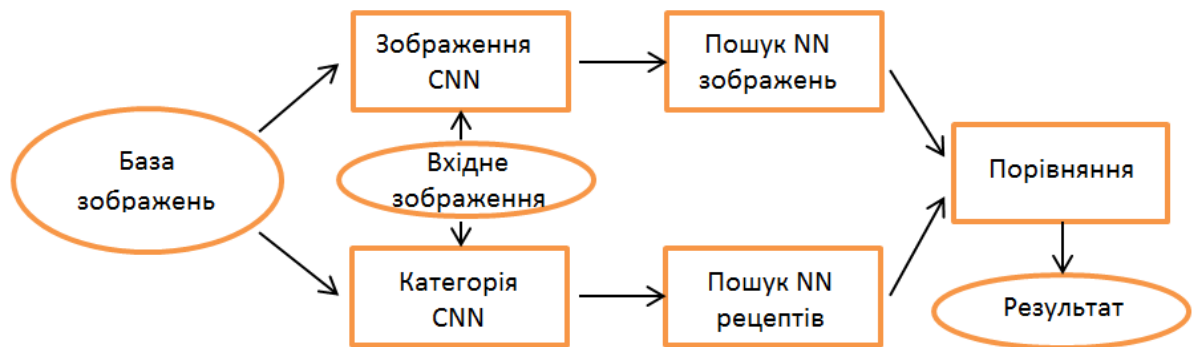


Рисунок 2.1 - Схема функціонування системи розпізнавання зображень

Загалом послідовний алгоритм за яким 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.

2.3 Тренування системи розпізнання

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, зазвичай розробляємо теплову карту, щоб отримати першу інформацію, які можливі функції цікаві. Карта дає зрозуміти, які значення співвідносяться з іншими значеннями.

Якщо видалити всі інгредієнти, які зустрічаються не один раз, залишаться унікальні. Для аналізу асоціацій інгредієнтів використовується алгоритм *argiori*. 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 п'ятірки.

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. спагетті (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:

спагетті карбонара олія

Тема №1:

салат із дині цикорій редька селера

Тема №2:

локшина китайська азійський 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 рисунку. 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 власних даних. Це економить 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 $n \times 4096$, де n - кількість вхідних зображень.

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



Рисунок 2.2 - Вектор 4096 обчислений від зображення тістечка

Тепер все, що потрібно зробити - це 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: гіперплощина LSH та 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 отриманої страви.

2.4 Визначення чутливості системи до якості зображення

Розробка алгоритму включає дослідження чутливості CNN до різних масштабів вихідного зображення, яке має десять варіантів із коефіцієнтом масштабу в межах від 0,6 до 1,5. Таким чином, десять форматів зображень повинні бути доставлені на вхід CNN для обробки.

Матриця відповіді сканування зображення нейронною мережею представлена на рисунку 3.

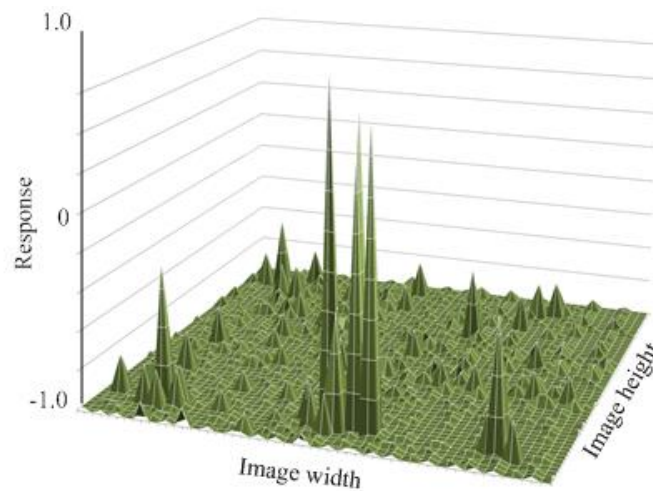


Рисунок 2.3 - Матриця відповіді CNN при скануванні зображень [3]

Відповіді CNN мають різні параметри в різних масштабах зображення. Це пояснюється тим, що візуальна структура особливостей об'єкта залежить від масштабу зображення, і CNN реагує на них по-різному.

Найефективніша локалізація таблички спостерігається на зображеннях з коефіцієнтами масштабу 0,8, 0,9, 1,0, 1,1, 1,2, як показано на рисунку, оскільки зображення в цих діапазонах менш схильне до спотворень і здебільшого відповідає вхідному розміру CNN.

Отже, масштабування зображення використовується для підвищення ефективності алгоритму локалізації, що дозволяє значно підвищити його продуктивність. Цей алгоритм включає наступні кроки.

Крок 1. Перетворення кольорового зображення у відтінки сірого.

Крок 2. Масштабування вихідного зображення з використанням діапазону масштабних коефіцієнтів 0,8 -1,2. Отримано набір із п'яти зображень різного формату.

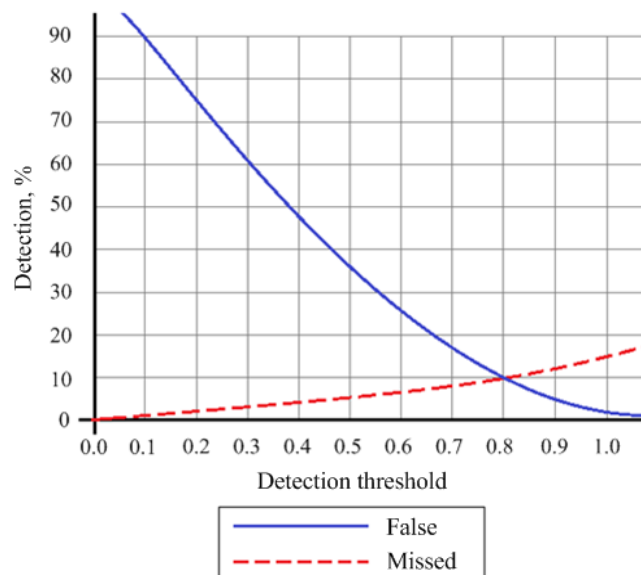
Крок 3. Попередня локалізація. Отримане зображення доставляється на перший вхід CNN. Він сканує зображення, щоб виявити ділянки з можливою присутністю об'єктів, тобто ділянки з відгуками CNN, що перевищують

запропонований поріг. Крок сканування дорівнює чотирьом пікселям. На цьому кроці можна виявити помилкові розташування об'єктів.

Кожне розташування об'єкта характеризується параметрами x , y , s , h , r , де x , y - координати центру розташування об'єкта; s - ширина розташування об'єкта; h - висота розташування об'єкта; r - реакція нейронної мережі на розташування об'єкта.

Крок 4. Остаточне розпізнання. Другий CNN сканує розташування об'єктів, виявлених на кроці 3. Крок сканування дорівнює одному пікселю. Відповіді CNN в межах локалізації повинні перевищувати запропонований поріг на всіх масштабах зображення. Деякі помилкові виявлення видаляються на цьому етапі.

Поріг виявлення (значення відгуку CNN, при якому виявляється розташування об'єкта в оброблюваній зоні) вибирається таким чином, щоб забезпечити максимум справжніх і мінімум помилкових виявлень номерних знаків зображення автомобіля. Експерименти показують, що 0,8 є оптимальним значенням для порогу виявлення. Результати експерименту представлені на рисунку 2.4.



Рисунк 2.4 - Помилкові та пропущені виявлення при різних значеннях порогу виявлення

Крок 5. Масштабування зображення відносно початкового розміру.

Крок 6. Порівняння параметри розташування об'єкта, а саме: координати та межі. Місцезнаходження із загальними параметрами вважаються істинними. Усі інші місця вважаються хибними та не враховуються.

Висновки до розділу 2

Отже 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.

Розділ 3

Розробка інформаційної технології розпізнавання зображень

3.1 Формування згорткової нейронної мережі

Спроби відтворити здатність біологічних нервових систем навчатися та виправляти помилки привели до створення штучних нейронних мереж. Штучні нейронні мережі являють собою сімейство моделей, побудованих за принципом організації та функціонування біологічних нейронних мереж – мереж нервових кліток живого організму.

Поняття штучної нейронної мережі було запропоновано ще в 1943 році У. Маккалоком і У. Питтсом у статті. Зокрема, ними була запропонована модель штучного нейрона.

Щоб відбити суть біологічних нейронних систем, штучний нейрон будується в такий спосіб. Він одержує вхідні сигнали (вихідні дані або вихідні сигнали інших нейронів нейронної мережі) через кілька вхідних каналів. Кожний вхідний сигнал проходить через з'єднання, що має певний вагу. З кожним нейроном зв'язане певне граничне значення. Обчислюється зважена сума входів, з неї віднімається граничне значення та у результаті виходить величина активації нейрона. Сигнал активації перетвориться за допомогою функції активації та у результаті виходить вихідний сигнал нейрона.

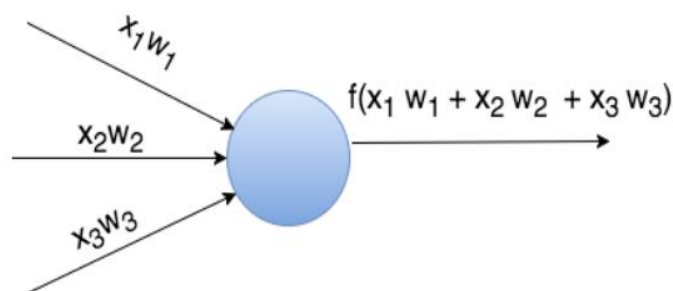


Рисунок 3.1 - Штучний нейрон [10]

x_i – вхідний сигнал

w_i – вага вхідного сигналу

$f(\cdot)$ – функція активації.

3.1 Функції активації

У даному розділі описані використовувані в роботі функції активації для нейронних мереж.

Сигмоїдна: $f(s) = \frac{1}{1+e^{-s}}$

Лінійна: $f(s) = s$

Позитивно лінійна: $f(s) = \max(0, s)$

Софтмакс (Softmax): $f(s)_j = \frac{e^{s_j}}{\sum_{k=1}^K e^{s_k}}$ для $j = 1, \dots, K$

3.2 Функція втрат

Уведемо позначення: X – множина описів об'єктів, Y – множина припустимих відповідей. Передбачається, що існує невідома цільова залежність – відображення $y^* : X \rightarrow Y$ значення якої відомі тільки на об'єктах кінцевої навчальної вибірки $X^m = \{(x_1, y_1), \dots, (x_m, y_m)\}$.

Вводиться функція втрат $L(y, y')$, що характеризує величину відхилення відповіді y від правильної відповіді $y' = y^*(x)$ на довільному об'єкті $x \in X$. Тоді емпіричний ризик – функціонал якості, що характеризує середню помилку на навчальній вибірці:

$$Q(x^m) = \frac{1}{m} \sum_{i=1}^m L(y_i, y^*(x_i)) \quad (3.1)$$

У процесі навчання нейронна мережа набудовує ваги W , мінімізуючи емпіричний ризик.

При розв'язку задачі багатокласової класифікації на виході нейронної мережі необхідно одержати ймовірність приналежності об'єкта кожному із класів. У цьому випадку в якості функції втрат звичайно використовується крос-ентропія

$$L_{y_i, y^*(x_i)} = - \sum_{j=1}^K y_{ij}^* \log y_{ij} \quad (3.2)$$

K - кількість міток класів у задачі.

У даній роботі для класифікації текстів за допомогою нейронних мереж використовується описана функція втрат.

З появою більших об'ємів даних і більших обчислювальних можливостей стали активно використовуватися нейронні мережі. Особливу популярність одержали згорткові нейронні мережі, архітектура яких була запропонована Яном Лекуном і націлена на ефективне розпізнавання зображень. Своя назва архітектура мережі одержала через наявність операції згортки, суть якої в тому, що кожний фрагмент зображення множиться на матрицю (ядро) згортки поелементно, а результат підсумовується та записується в аналогічну позицію вихідного зображення. В архітектуру мережі закладені апріорні знання із предметної області комп'ютерного зору: піксель зображення сильніше пов'язаний із сусіднім (локальна кореляція) і об'єкт на зображенні може зустрітися в будь-якій частині зображення.

Особлива увага згорткові нейронні мережі одержали після конкурсу Imagenet, який відбувся в жовтні 2012 року та був присвячений класифікації об'єктів на світлинах. У конкурсі було потрібно розпізнавання образів в 1000 категорій. Переможець даного конкурсу – Алекс Крижевський, використовуючи згорткову нейронну мережу, значно перевершив інших учасників.

Успіх застосування згорткових нейронних мереж до класифікації зображень привів до множини спроб використовувати даний метод до інших

задач. Останнім часом їх стали активно використовуватися для задачі класифікації текстів.

3.4 Архітектура згорткової нейронної мережі

Згорткова нейронна мережа звичайно являє собою чергування згорткових шарів (convolution layers), субдискретизуючих шарів (subsampling layers) і наявності повнозв'язних шарів (fully-connected layer) на виході. Усі три види шарів можуть чергуватися в довільному порядку.

У згортковому шарі нейрони, які використовують ті самі ваги, поєднуються в карти ознак (feature maps), а кожний нейрон карти ознак пов'язаний із частиною нейронів попереднього шару. При обчисленні мережі виходить, що кожний нейрон виконує згортку деякої області попереднього шару (обумовленої множиною нейронів, пов'язаних з даним нейроном).

Приклад архітектури згорткової нейронної мережі представлено на рисунку 3.2.

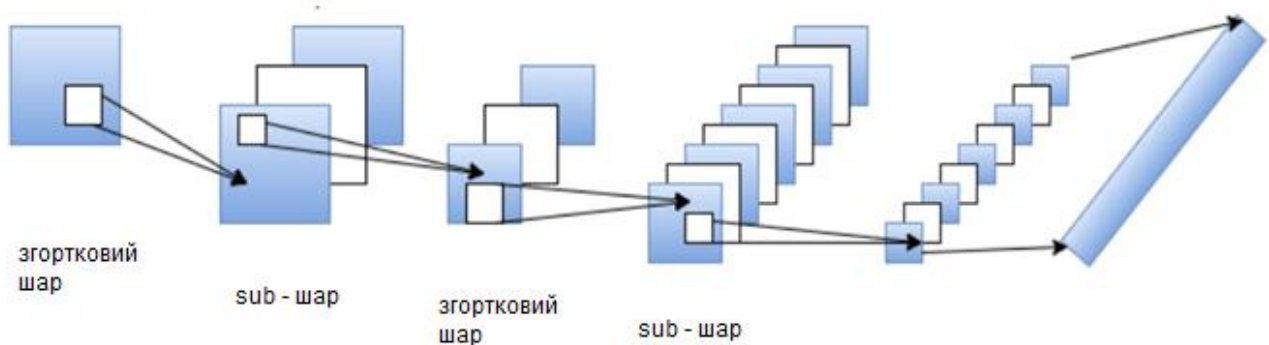


Рисунок 3.2 - Архітектура згорткової нейронної мережі [3]

3.5 Зв'язний шар

Шар у якому кожний нейрон з'єднаний з усіма нейронами на попередньому рівні, причому кожний зв'язок має свій ваговий коефіцієнт. На рисунку 3.3 показаний приклад повнозв'язного шару.

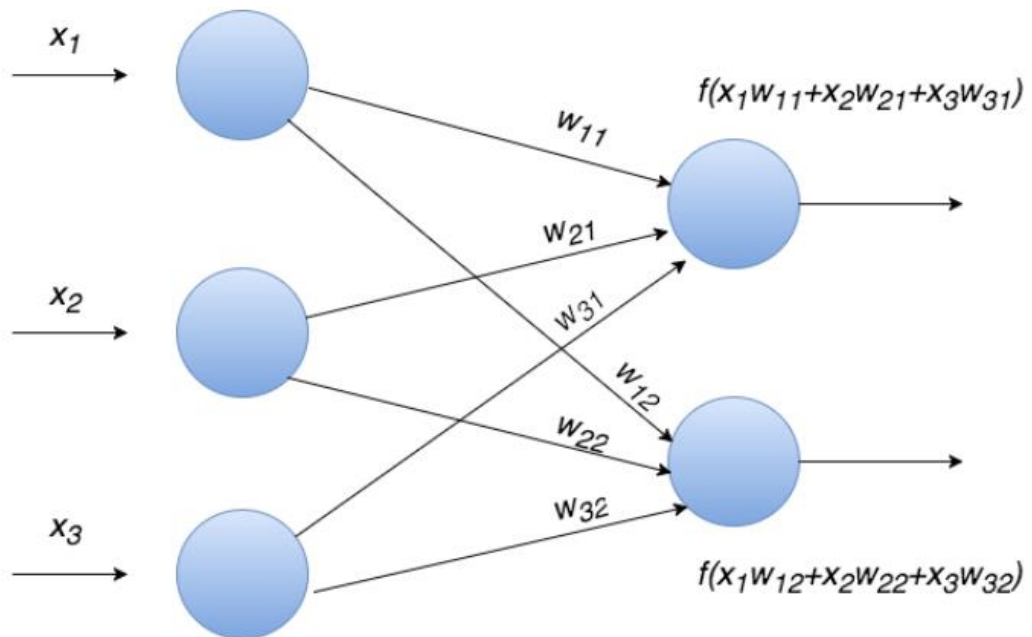


Рисунок 3.3 - Повнозв'язний шар [6]

w_{ij} – вагові коефіцієнти.

$f(\cdot)$ – функція активації.

3.6 Згортковий шар

На відміну від повнозв'язного, у згортковому шарі нейрон з'єднаний лише з обмеженою кількістю нейронів попереднього рівня, тобто згортковий шар аналогічний застосуванню операції згортки, де використовується лише матриця ваг невеликого розміру (ядро згортки), яку «рухають» по всьому оброблюваному шару.

Ще одна особливість згорткового шару в тому, що він небагато зменшує зображення за рахунок крайових ефектів.

На рисунку 3.4 показаний приклад згорткового шару з ядром згортки розміру 3×3 .

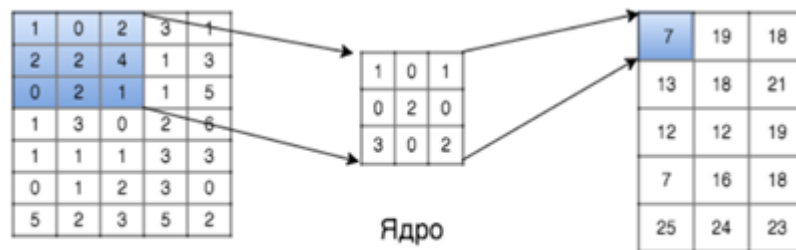


Рисунок 3.4 - Згортковий шар [4]

Шари цього типу виконують зменшення розмірності (звичайно в кілька раз). Це можна робити різними способами, але найчастіше використовується метод вибору максимального елемента (max-pooling) – уся карта ознак розділяється на гнізда, з яких вибираються максимальні за значенням.

На рисунку 3.5 показаний приклад субдискретизуючого шару з методом вибору максимального елемента.

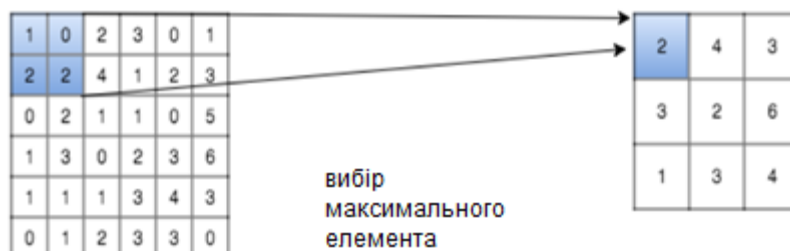


Рисунок 3.5 - Субдискретизуючий шар [11]

Dropout шар (шар регуляризації) спосіб боротьби з перенавчанням у нейронних мережах, навчання яких звичайно роблять стохастичним градієнтним спуском, випадково вибираючи деякі об'єкти з вибірки. Dropout-регуляризація полягає в зміні структури мережі: кожний нейрон викидається з деякою ймовірністю p . По такій прорідженій мережі проводиться навчання, для ваг, що залишилися, робиться градієнтний крок, після чого всі викинуті нейрони повертаються в нейромережу. Таким чином, на кожному кроці стохастичного градієнта набудовуємо одну з можливих 2^N архітектур мережі, де під архітектурою розуміємо структуру зв'язків між нейронами, а через N позначаємо сумарне число нейронів. При тестуванні нейромережі нейрони вже не викидаються, але вихід кожного нейрона множиться на $(1 - p)$ – завдяки цьому на виході нейрона будемо одержувати матоочікування його відповіді по всім 2^N архитектурам. Таким чином, навчену за допомогою dropout-регуляризації нейромережу можна розглядати як результат усереднення 2^N мереж.

На рисунку 3.6 представлена суть методу Dropout.

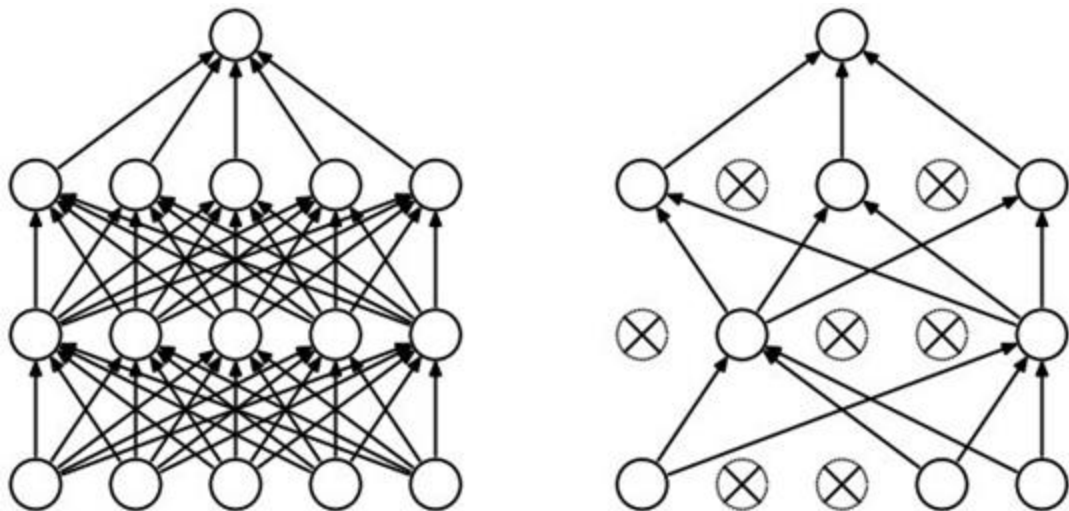


Рисунок 3.6 - Графічне представлення методу Dropout. Ліворуч – нейронна мережа до того, як до неї застосували Dropout, праворуч – та ж мережа після Dropout. [5]

3.7 Моделі використання згорткової нейронної мережі

У даному розділі будуть описані основні підходи використання згорткових нейронних мереж для задачі класифікації текстів.

Назвемо алфавітом упорядкований набір символів. Нехай обраний алфавіт складається з m символів. Кожний символ алфавіту в тексті закодовано за допомогою $1-m$ кодування (кожному символу буде співставлений вектор довжини m елемент якого дорівнює одиниці, у позиції рівній порядковому номеру символу в алфавіті, а нулю у всіх інших позиціях.) Якщо в тексті зустрінеться символ, який не ввійшов в алфавіт, то необхідно закодувати його вектором довжини m , що складається із одних нулів. З тексту вибираються перші l символів. Параметр l повинен бути більшим, щоб у перших l символах утримувалося досить інформації для визначення класу всього тексту.

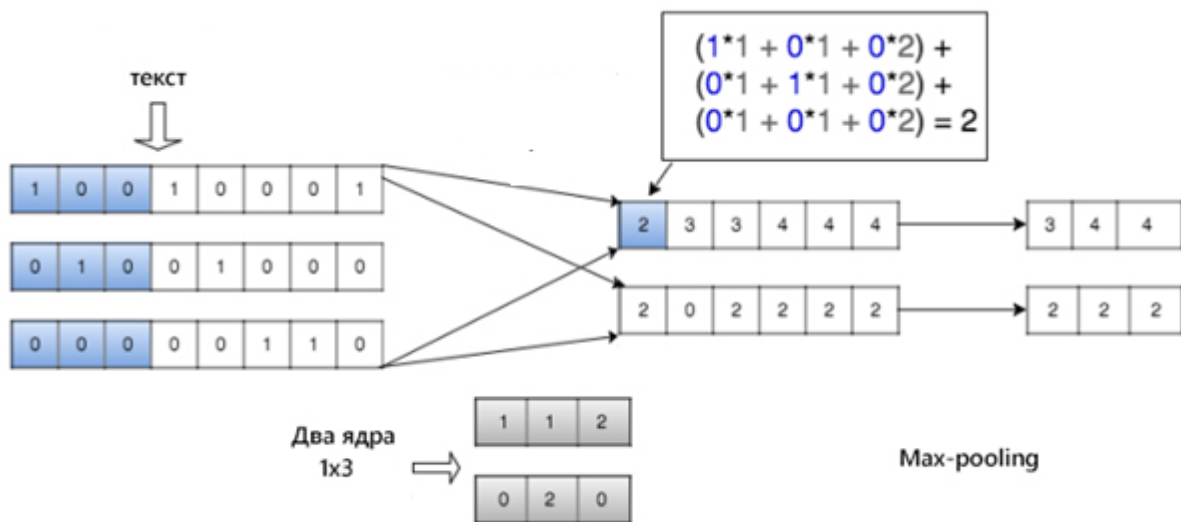


Рисунок 3.7 - Посимвольний підхід [5]

Далі отримані вектори складаються в матрицю розміру $m \times l$, у якій кожний стовпець буде мати не більш однієї одиниці. Кожний рядок отриманої матриці використовується як окрема карта ознак. На вхід згорткової нейронної

мережі подається m карт ознак розміру $l \times l$ аналогічно зображенню. Архітектуру мережі необхідно вибирати виходячи із задачі. На рисунку 3.8 наведений приклад посимвольного підходу для $l=6, m=3$. У прикладі показано один згортковий і один субдискретизуючий шар.

Опишемо формально даний підхід.

Нехай x_i – вектор i -го символу в тексті.

$$x_{1:l} = x_1 \oplus x_2 \oplus \dots \oplus x_l \quad (3.3)$$

Тут \oplus операція об'єднання векторів.

$$c_i = cw \cdot x_{i:i+h-1} + b \quad (3.4)$$

$$c = [c_1; c_2; \dots; c_{n-h+1}]$$

f – функція активації нейронної мережі

b – константа

Max-pooling шар:

$$\hat{c} = \max(c) \quad (3.5)$$

Dropout шар:

$$y = w(z \circ r) + b \quad (3.6)$$

де \circ – посимвольне множення.

r – вектор, що складається з нулів і одиниць.

У даному підході кожному слову в тексті співставляється вектор фіксованої довжини, потім з отриманих векторів для кожного об'єкта вибірки складається матриця, яка аналогічно зображенням подається на вхід згорткової

нейронної мережі. На рисунку 3.8 наведений приклад згорткової нейронної мережі.

Використання лише одного алгоритму часто виявляється неефективним у вирішенні складних задач. Отже, доцільніше реалізувати на основі адаптивного підсилення алгоритмічну композицію, що включає кілька класифікаторів CNN, кожен з яких компенсує помилки, отримані з попередніх ітерацій.

Архітектура першого CNN простіша і призначена для попереднього визначення вмісту зображення.

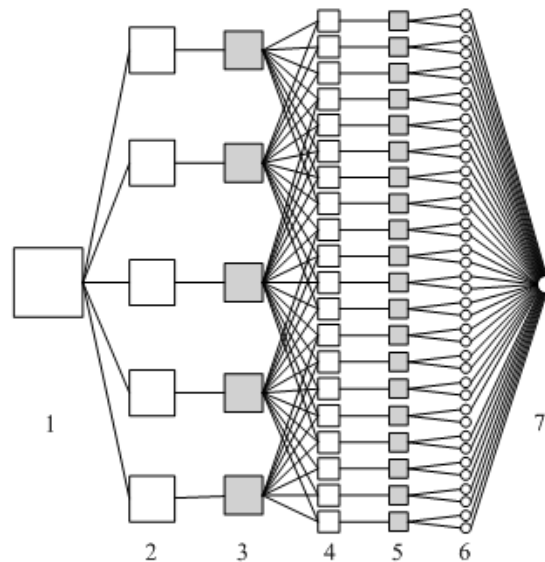


Рисунок 3.8 - Шари в згортковій нейронній мережі, призначені для попередньої локалізації: 1) вхідний шар; 2, 4) згорткові шари; 3, 5) субопробірні шари; 6) шар з нейронами; 7) вихідний рівень [5]

Враховуючи все, запропонований алгоритм виявлення розроблений у формі алгоритмічної композиції, що включає два CNN, що мають різні конфігурації. Перший CNN виконує попередню локалізацію таблички, тоді як другий забезпечує його остаточне виявлення, щоб компенсувати помилки класифікації, отримані першим CNN.

Вхідний рівень містить 22×60 нейрони. Другий згортковий шар складається з п'яти площин, що містять 20×52 нейрони. Розмір маски згортки становить 3×9 нейрони.

Третій субвибірка має таку ж кількість площин, як і попередній шар згортки. Кожна площина з'єднана лише з однією площиною попереднього шару і містить 10×26 нейронів, що вдвічі менше розміру попереднього шару. Розмір підвибірка маска 2×2 нейрони.

Четвертий згортковий шар складається з 20 площин 8×18 нейроформи. Кожна площина перехресно пов'язана з відповідними трьома площинами попереднього шару для збору наборів даних про варіації. Розмір маски згортки становить 2×2 нейрони.

П'ятий підвдбір шар складається з 20 площин 4×9 нейрони за розміром. Кожна площина пов'язана лише з однією площиною попереднього шару. Розмір підвибірки маски 2×2 нейрони.

Для CNN перші п'ять шарів служать для вилучення. Наступні два шари є для класифікація зразків.

Шостий шар складається з 40 нейронів. Кожен нейрон повністю пов'язаний з кожним нейроном пари відповідних площин, що містяться в попередньому шарі.

Сьомий шар складається з одного нейрона, який повністю пов'язаний з усіма нейронами попереднього шару. Вихідні значення цього нейрона знаходяться в діапазоні від -1 до $+1$.

Архітектура другого CNN є більш складною, оскільки вона має більшу кількість площин і нейронів і призначена для остаточної локалізації.

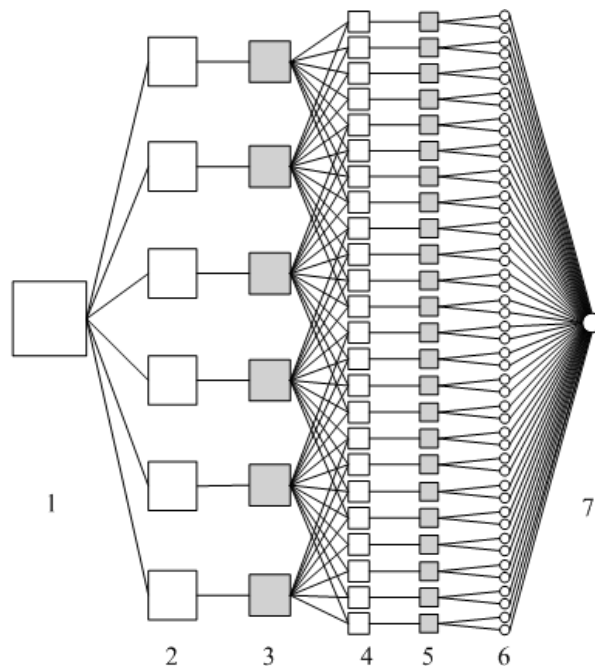


Рисунок 3.9 - Шари в згортковій нейронній мережі, призначені для остаточної локалізації: 1) вхідний шар; 2, 4) згорткові шари; 3, 5) субопробірни шари; 6) шар з нейронами; 7) вихідний рівень [6]

У розвинених CNN функціонування нейронів у згортковому шарі визначається як

$$y_k^{(i,j)} = b_k + \sum_{s=1}^K \sum_{t=1}^K w_{k,s,t} x^{((i-1)+s, (j+t))}, \quad (3.7)$$

де $y_k^{(i,j)}$ - нейрон k -ї площини згорткового шару; b_k - нейронне зміщення k -ї площини; K - розмір рецептивного поля нейрона; $w_{k,s,t}$ - матриця синаптичного коефіцієнта; x - вихід нейрону з попереднього шару.

Функціонування нейронів у шарі субвиборів визначається як

$$y_k^{(i,j)} = b_k + \frac{1}{4} w_k \sum_{s=1}^2 \sum_{t=1}^2 x^{((i,j)+s, (i,j))}. \quad (3.8)$$

Розмір площини в шарі згортки обчислюється як:

$$s_k = s_p - N + 1, \quad h_k = h_p - M + 1, \quad (3.9)$$

де s_k, h_k - це ширина та висота площини згортки відповідно; s_p, h_p - це ширина та висота попередньої площини шару; N, M - це ширина і висота рецептивного поля.

Розмір площини в шарі підпроби обчислюється як:

$$s_k = \frac{s_p}{2}, \quad h_k = \frac{h_p}{2}, \quad (3.10)$$

де s_k, h_k - це ширина та висота площини згортки відповідно; s_p, h_p - це ширина та висота попередньої площини шару.

Висновки до розділу 3

У цій роботі запропоновано алгоритм локалізації або виявлення забраження. Цей алгоритм представлений алгоритмічною композицією, що складається з двох CNN, що мають різні конфігурації. Перший CNN забезпечує попередню локалізацію, тоді як другий забезпечує остаточне розпізнання. Таким чином, другий CNN компенсує класифікаційні помилки, отримані першим CNN. Цей алгоритм показує справжню ймовірність виявлення, що дорівнює 82 відсотка у найкращих спробах.

Експерименти показали, що реалізований алгоритм забезпечує високу ефективність.. Запропонований алгоритм можна успішно оптимізувати, а потім застосувати до виявлення інших об'єктів у зображеннях та відеопослідовностях.

Розділ 4

Дослідження роботи системи розпізнавання

4.1 Трансформування даних

Як і будь-яка інша область впливає на благополуччя людини в тій чи іншій мірі, як харчування. Щодня незліченна кількість харчових знімків публікується від користувачів у соціальних мережах; від першого домашнього торта до страви високої кухні, радість розділяється на випадок, якщо страва буде успішно приготовлена.

Це факт, що незалежно від того, наскільки страва відрізняється одна від одшої, хорошу їжу цінують всі.

Досягнення в класифікації окремих інгредієнтів приготування досить рідкісні. Проблема полягає в тому, що майже немає загальнодоступних відредагованих записів. У цій роботі йдеться про проблему автоматизованого розпізнавання сфотографованій кулінарної страви і подальшого виходу відповідного рецепта. Відмінність між складністю обраної проблеми і попередніми проблемами класифікації під наглядом полягає в тому, що існують великі перекриття в харчових стравах (наприклад, висока внутрішньо-класна схожість), так як страви різних категорій можуть виглядати дуже схожими тільки з точки зору інформації про зображення.

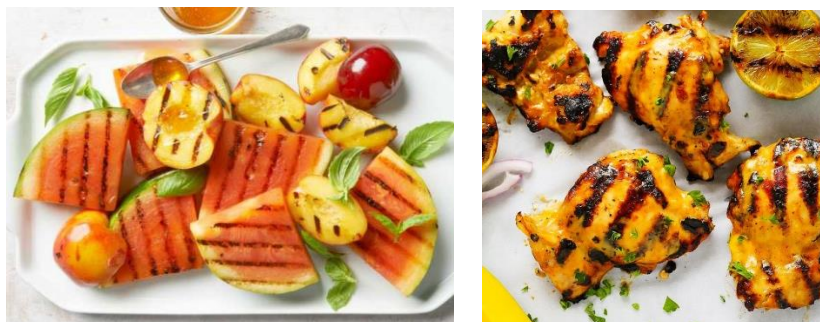


Рисунок 4.1 – Зображення різних блюд

Зображення схоже, але не схожі за змістом та рецептом приготування, та взагалі за смаковими якостями, тобто вони зовсім різні.

Згідно з нинішнім станом, найбільший німецькомовний набір даних з більш ніж 300 000 рецептів буде зібраний і проаналізований. Потім буде представлений розроблений метод: поєднання розпізнавання об'єктів або визнання кухонного суду за допомогою Convolutional Neural Networks (короткий CNN) і пошук найближчих сусідів (Next-Neighbor Classification) в записі понад 800 000 зображень. Ця комбінація допомагає знайти правильний рецепт більш імовірно, оскільки топ-5 категорій CNN порівнюються з наступною сусідньою категорією з рейтинговою кореляцією. Rank кореляція на основі підходів виміряє ймовірність двох пунктів, будучи в одному порядку в двох рейтингових списках. Математично Кендалл Тау обчислюється як

$$\tau = P(C) - P(D) = \frac{C}{N} - \frac{D}{N} = \frac{C-D}{N} \quad (4.1)$$

де: N - кагальна кількість пар;

C - кількість конкордантних пар;

D -кількість дискордантних пар.

Для кожного рецепта W є K кількість фотографій. Для кожного з цих зображень вектори функції витягуються з попередньо навченої Convolution Neural Network, навченої на 1000 категоріях в конкурсі розпізнавання зображень ILSVRC 2014 з мільйонами зображень. Вектори функцій формують внутрішнє представлення зображення в останньому повністю з'єднаному шарі перед 1000-категорією Softmax Layer, який був видалений заздалегідь. Ці вектори функцій потім вимірно скорочуються PCA (основний аналіз компонентів) з матриці N x 4096 до матриці N x 512. В результаті можна вибрати топ-5 зображень з найменшою відстанню евкліда до вхідного зображення (Приблизний

найближчий сусід), тобто топ-5 оптичних, просто з інформації про зображення, аналогічні зображення до вхідного зображення.

Крім того, CNN навчений S рядом категорій з фотографіями W рецептів. S було визначено динамічно за допомогою моделювання тем та семантичного аналізу назв рецептів. В результаті отримуємо для кожної категорії ймовірність того, до якої може належати вхідне зображення.

Категорії top-k від CNN порівнюються з категоріями з верхнього k оптично схожих зображень з кореляцією Кендалл Тау.

Схема візуалізації методу виглядає наступним способом:

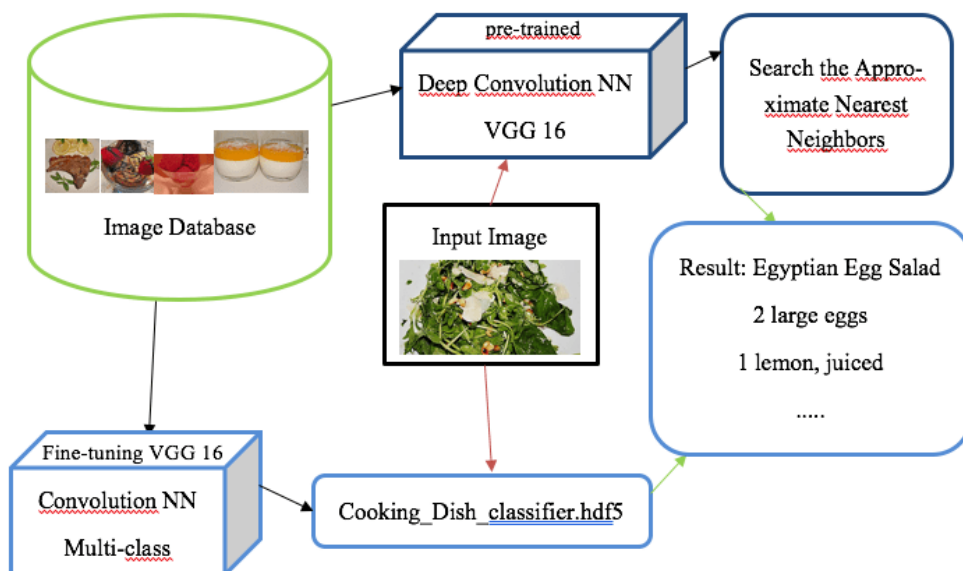


Рисунок 4.2 – Схема підходу

- 1 | — Підготовка
 - | — Очищення даних
 - | — Збільшення даних
- 2 | — Аналіз та візуалізація даних, розділені дані (Прохідний, Дійсний, Тест)
- 3 | — Тематичне моделювання

- | └─ Латентний Діріхте розподіл (LDA)
- | └─ Не негативні матричні факторизації
- 4 | ── Функція
 - | └─ k-найближчі сусіди
 - | └─ t-SNE візуалізація
- 5 | ── Трансферне навчання CNN (Convolutional Neural Network)
 - | └─ AlexNet, VGG, ResNet, GoogLeNet
- 6 └─ Розгортання додатку

4.2 Збір і підготовка даних

Для того, щоб мати можливість навчити модель взагалі, потрібно достатньо даних (так зване збільшення даних і точне налаштування попередньо навчених моделей можна використовувати як засіб). Тільки через таку кількість даних узагальнення навчального набору може бути постійно збільшено до деякої міри і висока точність може бути досягнута в тестовому наборі. Перша частина стосується збору даних, аналізу та візуалізації функцій та їх взаємозв'язків. У нас немає кращих алгоритмів, у нас просто більше даних.

Без винятку якість і кількість набору даних не є незначною. Саме тому найбільша в Європі кулінарна платформа буде зібрана: кожен рецепт, загалом, 316'756 рецептів (станом на 2020 року), завантажуються в цілому з 879'620 зображень. Важливо не продовжувати занадто швидко при завантаженні і захищати сервери від занадто великої кількості запитів, так як в іншому випадку заборона власної IP-адреси зробить збір даних більш важким.

Більше даних призводить до більших вимірів, але більше вимірів не обов'язково призводять до кращої моделі та її представлення. Відхилення закономірностей у наборі даних, які порушують навчання, може бути ненавмисно підсилене більшою кількістю вимірів, узагальнення та вивчення запису даних

погіршується для нейронної мережі, зменшується співвідношення сигналу до шуму.

При виконанні зібрані даних веб-сайту важливо поважати файл robots.txt. Деякі адміністратори не хочуть відвідування від ботів до конкретних каталогів.

```
User-agent: * # directed to all spiders, not just scooters
```

```
Disallow: / cgi-bin
```

```
Disallow: / stats
```

```
Disallow: / pictures / photo albums /
```

```
Disallow: / forumuploads /
```

```
Disallow: / pictures / user /
```

```
Disallow: / user /
```

```
Disallow: / avatar /
```

```
Disallow: / cms /
```

```
Disallow: / products /
```

```
Disallow: / how2videos /
```

Перераховані каталоги, які нас не цікавлять, тому можна продовжувати. Тим не менш, такі заходи, як випадкові заголовки і досить великі паузи між окремими запитами рекомендується, щоб уникнути можливої заборони з веб-сайту.

```
def _get_html(url):
    page = ''
    while page == '':
        try:
            page = requests.get(url, headers=random_headers())
        except:
            print('Connection refused')
            time.sleep(10)
            continue
    return page.text
```

```

def _get_total_pages(html):
    soup = BeautifulSoup(html, 'lxml')
    total_pages = soup.find('div', class_='ck-pagination qa-
pagination').find('a', class_='qa-pagination-pagelink-last').text
    return int(total_pages)

html_text_total_pages = _get_html(category_url)
total_pages = _get_total_pages(html_text_total_pages)
print('Total pages: ', total_pages)Total pages: 10560

```

Наступним важливим кроком є вибір функцій для усінення недоліків неважливих даних. Підготовка необроблених даних для нейронної мережі є звичайним явищем на практиці. У першому пропуску буде завантажено назву рецепта, програму для рецепта, кількість оцінок, рівень складності, час приготування та дату публікації. У другому пропуску буде надруковано список складових, текст рецепта, всі зображення і кількість разів, протягом яких було надруковано рецепт. За допомогою цих функцій запис даних можна дуже добре описати і допомагає отримати сильне розуміння набору даних, що важливо для вибору алгоритмів.

Такі дані, як назва рецепта, оцінка, дата вивантаження рецепта тощо, зберігаються у файлі csv. Якщо у рецепті є зображення, мініатюру буде поміщено до search_thumbnails файла. Зробимо використання багатопотоковості, щоб забезпечити коротший час завантаження.

```

def scrap_main(url):
    print('Current url: ', url)
    html = _get_html(url)
    _get_front_page(html)
    #sleep(randint(1, 2))start_time = time()
with Pool(15) as p:
    p.map(scrap_main, url_list)
print("--- %s seconds ---" % (time() - start_time))

```

Далі потрібно отримати список інгредієнтів, приготування, теги і всі зображення кожного рецепта.

```
def write_recipe_details(data):
    dpath = DATAST_FOLDER + DFILE_NAME
    with open(dpath, 'a', newline='') as f:
        writer = csv.writer(f)
        try:
            writer.writerow((data['link'],
                             data['ingredients'],
                             data['zubereitung'],
                             data['tags'],
                             data['gedruckt:'],
                             data['n_pics']
                             #data['reviews'],
                             #data['gespeichert:'],
                             #data['Freischaltung:'],
                             #data['author_registration_date'],
                             #data['author_reviews']
                             ))
        except:
            writer.writerow('')
```

Якщо все пройшло гладко з завантаженням, дані виглядають наступним чином:

Всього 879'620 зображень (35 ГБ) 316'756 рецептів - 3 яких 189'969 містять одну або кілька фотографій - 3 яких 107052 рецепти містять більше 2 зображень - 126'787 не містять зображення.

4.3 Аналіз даних і візуалізація

Для того, щоб отримати перше враження, зазвичай створюємо карту даних, щоб отримати перші ідеї, які можливі функції даних цікаві.

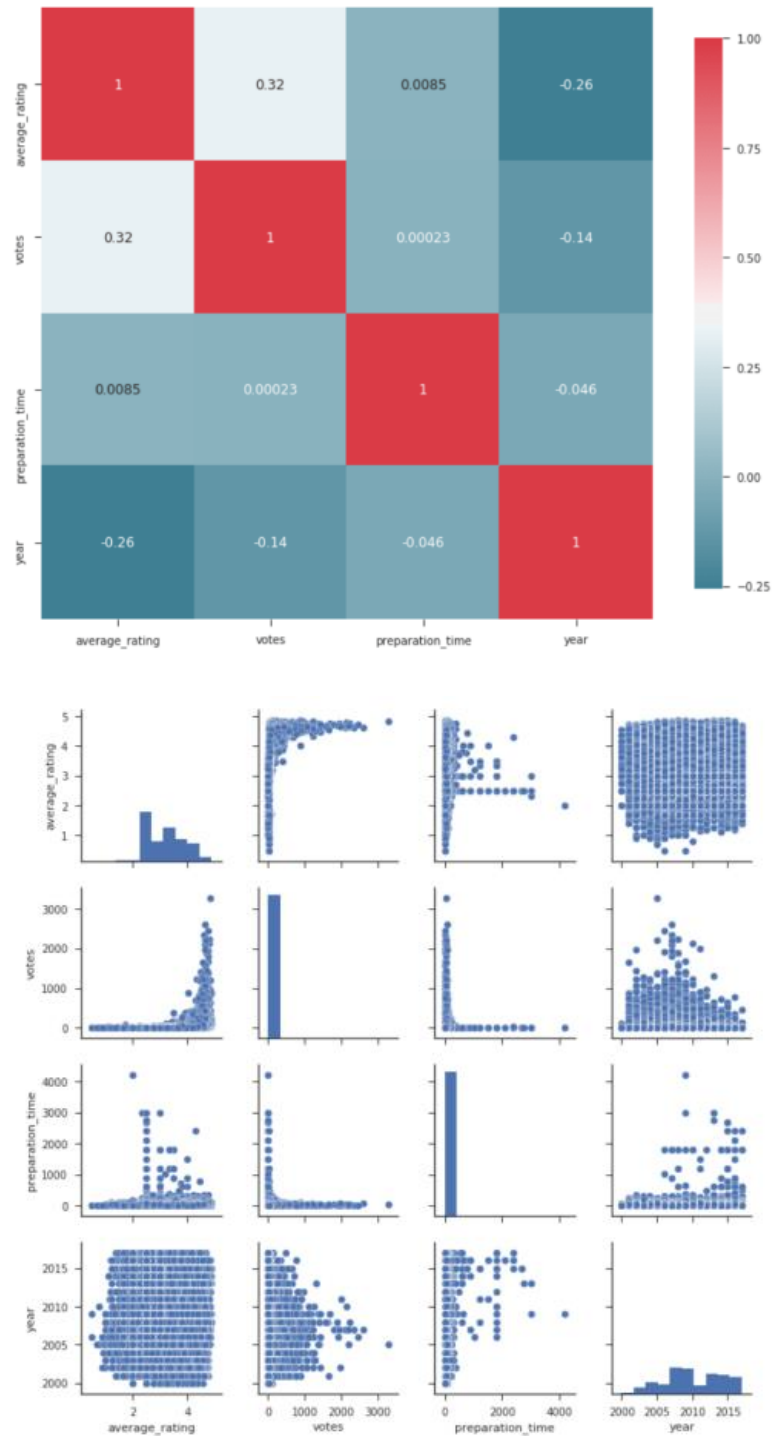


Рисунок 4.3 –Карта даних

Теплова карта дає уявлення про те, які цінності співвідносяться з іншими значеннями.

Найвища кореляція має голоси і середній рейтинг. На рисунку 4.3 показано сюжет пари на 1-му стовпчику, 2-му рядку, і виділяється, що чим вище кількість оцінок, тим кращий рейтинг рецепта. Також цікавим є порівняння між часом підготовки та кількістю оцінок. Більшість відгуків засновані на рецептах з коротким часом приготування. Схоже корпус віддає перевагу легким рецептам. Інша ідея полягає в порівнянні кількості нових завантажених рецептів на рік.

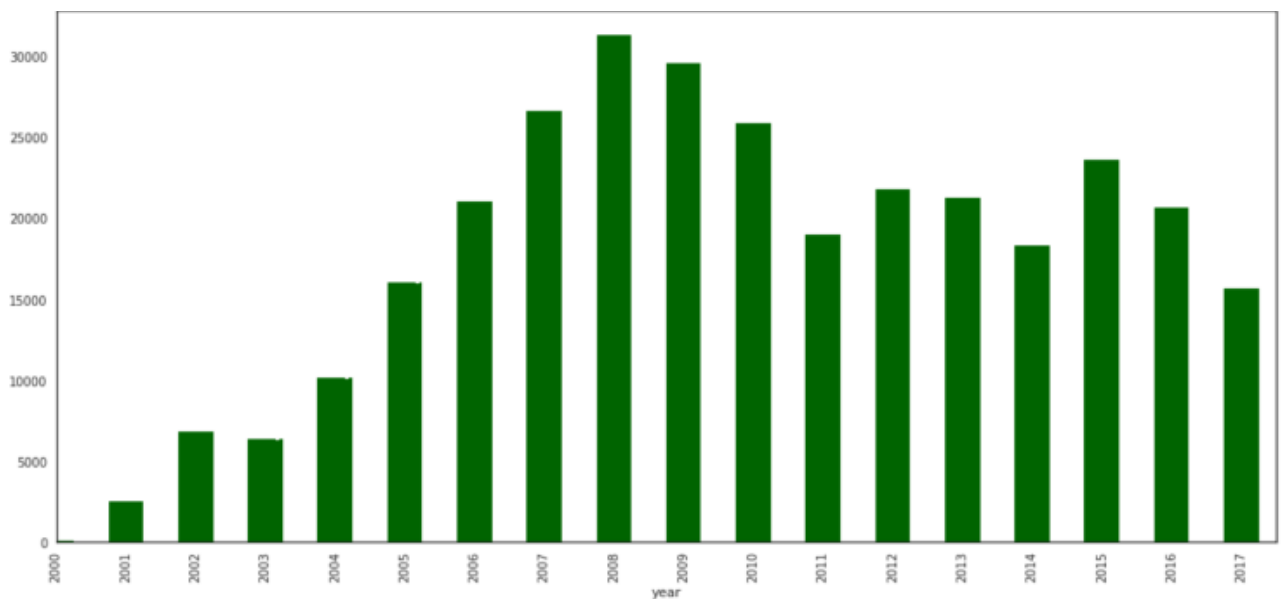


Рисунок 4.4 – Завантаження рецептів по роках

У 2008-2009 роках він має помітно найбільше завантажень на рік. Швидкий пошук пошуку в Інтернеті показує, що в 2008 році криза цін на продукти харчування переважала.

Порівняння кривих (нижній графік) показує, що існувала пряма кореляція між зростанням світових цін і постачанням рецептів. Гіпотеза полягає в тому, що

попит виріс за рецептами, тому що один залишився вдома і приготував для себе і своєї сім'ї, щоб заощадити бюджет і звести кінці з кінцями якомога більше.

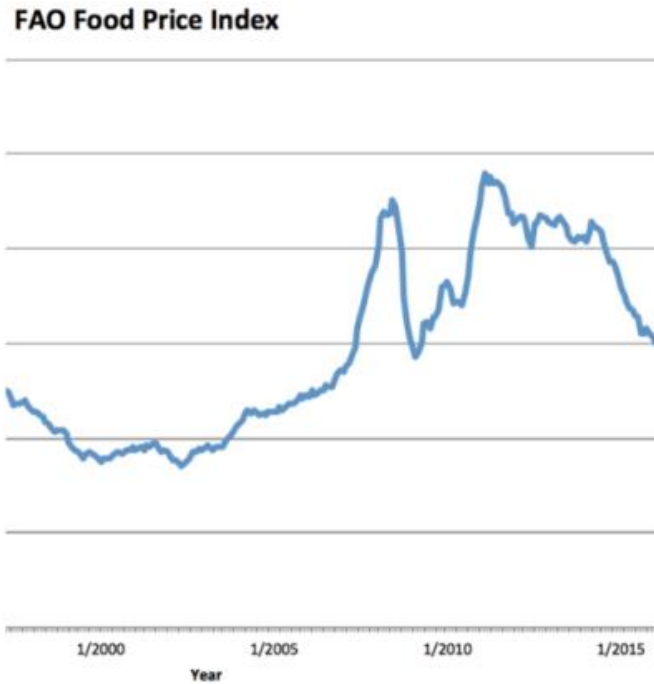


Рисунок 4.5 – Індекс цін

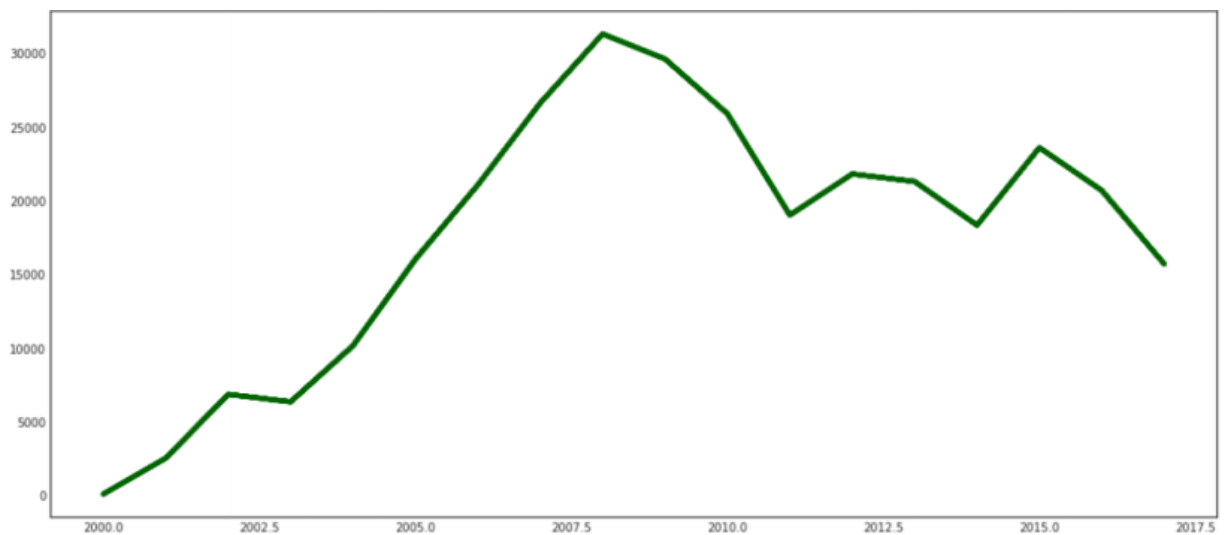


Рисунок 4.6 – Кількість завантажених рецептів на рік.

В цілому 316'755 рецепти поділяють 3'248'846 інгредієнтів. Якщо видалити всі інгредієнти, які використовуються більше одного разу, є 63'588 унікальних інгредієнтів. Це забезпечує частоту того, які інгредієнти в поєднанні з іншими інгредієнтами використовуються в цілому.

support	itemsets		
0.601187	[Salz]	0.227850	[Salz, Zwiebel]
0.432675	[Pfeffer]	0.205400	[Pfeffer, Zwiebel]
0.402075	[Pfeffer, Salz]	0.194013	[Pfeffer, Salz, Zwiebel]
0.357837	[Zucker]	0.185262	[Ei, Mehl]
0.301750	[Ei]	0.177787	[Ei, Zucker]
0.273625	[Mehl]	0.177575	[Mehl, Zucker]
0.272175	[Butter]	0.169300	[Ei, Salz]
0.269962	[Zwiebel]	0.167313	[Öl]

Рисунок 4.7 – Ліворуч - топ-8, а праворуч - 9-16 найкращих інгредієнтів з найвищим рівнем використання

Лідером інгредієнтів є сіль з 60-відсотковим представленням у всіх рецептах. На третьому місці можна побачити перший кортеж, поєднання двох інгредієнтів, а саме перцю і солі з трохи більше 40 відсотків вони на сьогоднішній день є найпоширенішою парою. Найпоширеніші трійнята, чотиримісні і навіть п'ятикутники можна знайти у відповідному поєднанні рецепту.

4.4 Моделювання тем

Метою цієї процедури є поділ всіх назв рецептів на n-категорії. Для контрольованої проблеми класифікації повинні забезпечити нейронну мережу зображеннями з мітками. Тільки з цими етикетками навчання стає можливим.

Проблема полягає в тому корпус не класифікує свої фотографії. Тому повинні зробити це самостійно. Можливі процедури для розділення назв рецептів 316'755 наведено нижче.

Візьмемо наступний приклад:

Піца з грибами.

Фарширований перець з горохом і тунцем.

Піца з морепродуктами.

Паприка з горохом.

Чотири назви рецептів, наведених вище, слід розділити на n категорій. Очевидно, що 1-й і 3-й рецепт потрібно бути в одній категорії під назвою піца. 2-й і 4-й також можна розділити на нову категорію за рахунок гороху.

4.5 Латентний розміщення Діріхте

(LDA - Latent Dirichlet allocation) це модель машинного навчання та пошуку інформації, яка може пояснити спостереження через неявні групи, таким чином визначаючи подібність між певними частинами даних. Наприклад, якщо в спостереженнях є слова, зібрані в документах, то зазначається, що кожен документ є сумішшю невеликої кількості тем і що поява кожного слова пов'язана з однією з тем документа. LDA є одним з методів тематичного моделювання і вперше був представлений як модель графіка для виявлення тем Девіда Блей, Ендрю Юна і Майкла Джордана в 2003 році.

У LDA кожен документ можна розглядати як набір різних тем. Такий підхід схожий на латентно-семантичний аналіз (pLSA) з різницею, яку LDA припускає, що розподіл об'єкта має як апіорі розподіл Dirichle. На практиці результатом є більш правильний набір тем.

LDA - це модель ймовірності, яка передбачає, що кожне ім'я можна призначити темі. По-перше, імена повинні бути очищені, тобто stop слова видаляються і слова зводиться до їх кореня. Чистий словник служить вхідним.

```

de_stop = get_stop_words('german')
s_stemmer = SnowballStemmer('german')
tokenizer = RegexpTokenizer(r'\w+')
final_names = []

for recipe_name in twentyeighth_iter:
    raw = recipe_name.lower()
    tokens = tokenizer.tokenize(raw)
    stop_t = [recipe_name for recipe_name in tokens if not recipe_name in
de_stop and not recipe_name in filter_words_]
    stem_t = [i for i in stop_t if len(i)>1]
    if len(stem_t)==0: final_names.append(['error'])
    else: final_names.append(stem_t)

print('20 Cleaned Recipe names example: \n >>>')
pprint(final_names[:20])20 Cleaned Recipe names example:
>>>
[['bratapfel', 'rotkohl'],
 ['frühstückswolke'],
 ['deichgrafensalat'],
 ['geschichteter', 'kohl'],
 ['rinderlendenragout'],
 ['blaukraut'],
 ['sauerbraten'],
 ['punschtorte'],
 ['oberländer'],
 ['mcmoes', 'pasta'],
 ['geschnetzeltes'],
 ['ahorn', 'bacon', 'butter'],
 ['endiviensalat'],
 ['rote', 'linsen', 'gemüse'],]
 ['kotelett', 'gratin'],

```

```
['rotkohl'],
['remouladensauce'],
['nudeln'],
['kohlsuppe'],
['gemüse', 'hackfleischauflauf']
```

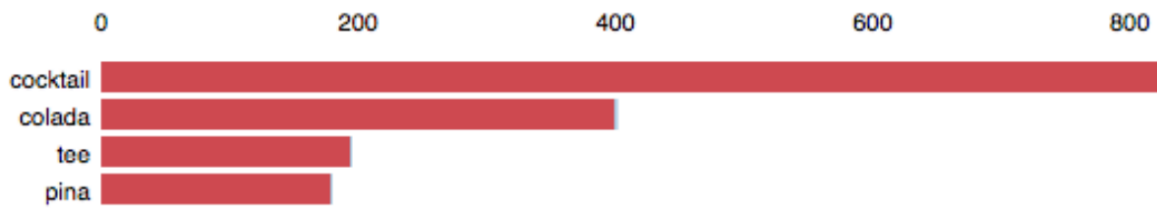


Рисунок 4.8 – Найбільш релевантні теми

300 тем були встановлені як умова. Модель для теми 89 дає хороші результати: напої виявляються та узагальнюються.

Заради простоти точне математичне визначення не обговорюється. В результаті, є список ймовірностей того, наскільки певна модель в тому, що вона буде відповідати темі. Приклад: '0.363 *' запечений '+' 0.165 *' пряний '+' 0.124 *' літо '+' 0.006 *" спеції "+" 0.004 *" вівсяне печиво ".

4.6 Не негативна матрична факторизація

Першим кроком є обчислення tf-idf. Це являє собою не більше, ніж важливість слова в назві рецепта, враховуючи важливість у всьому корпусі тексту.

Чотири найважливіші слова:

1. салат (2935.18)
2. спагетті (2429.36)
3. пряності (2196.21)
4. торт (1970,08)

Алгоритм NMF приймає як введення tf-idf і одночасно виконує зменшення розмірів і кластеризації. Ці зусилля дають хороші результати, як заявлено нижче для перших 4 тем:

Тема #0:

спагетті карбонара, олія

Тема #1: салат

кукурудзяні чікоре, булгур, редька, селера.

Тема #2:

локшина китайська, азіатський вок, удон, базилік.

Тема #3:

кекси, чорничний фундук, журавлина соковита, вівсянка

Результат можна візуалізувати за допомогою t-SNE. Важливо, щоб запис з декількома вимірами зводиться до 2D, що дозволяє знайти координату для кожної назви рецепта.

Через складність і пов'язаний з цим час розрахунку t-SNE було взято підмножину в 50 000.

Нейронні мережі працюють, відображаючи модель людського мозку. Ідея полягає в тому, що вона вчиться на своїх помилках, поступово підлаштовують вагу нейрона, щоб адаптуватися до даних.

За допомогою CNN інформація про зображення спочатку підсумовується, щоб зменшити кількість параметрів. припускаємо, що перші шари в CNN визнають грубі структури на картинці. Тим далі переходимо до останнього шару Softmax, тим тоншими стають вивчені функції.



Рисунок 4.9 – Візуалізація рецептів

Можемо скористатися цим і приймаємо попередньо навчені CNN, які були навчені мільйонами фотографій і видалити останні шари, щоб навчити їх власним даним. Це економить мільйони параметрів і, таким чином, скорочує обчислювальний час. CNN обраний тут VGG-16, який був підготовлений в класифікації конкурсу 2014 на 1000 категорій.

Якщо видалити останній шар, то отримаємо особливість селекції другого-останнього шару. Це формує матрицю $n \times 4096$, де n — це кількість вхідних зображень.

```
features = []
for image_path in tqdm(images):
    img, x = get_image(image_path);
```

```
feat = feat_extractor.predict(x) [0]
features.append(feat)
```

Дозволяємо VGG-16 розрахувати вектор для кожного зображення. Цей вектор є, так би мовити, відбитком малюнка: внутрішнє представлення буде нейронна мережа.

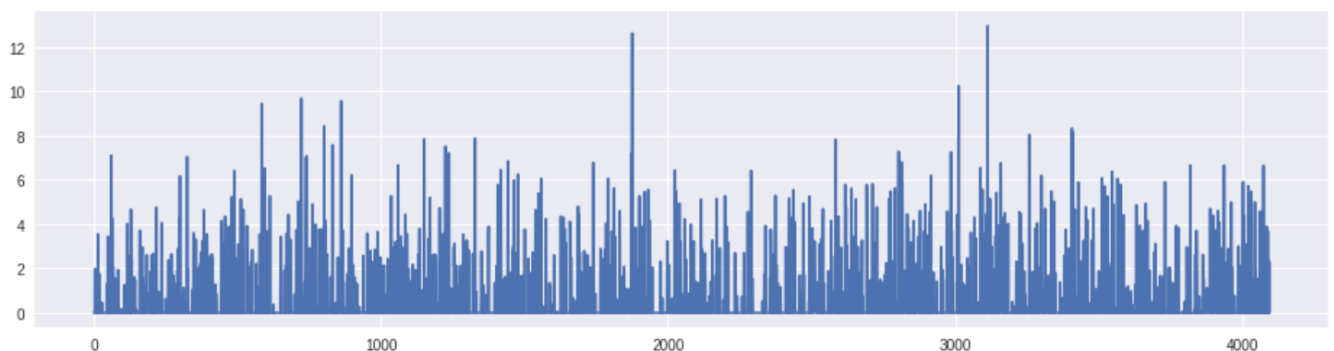


Рисунок 4.10 – Побудова вектора 4096

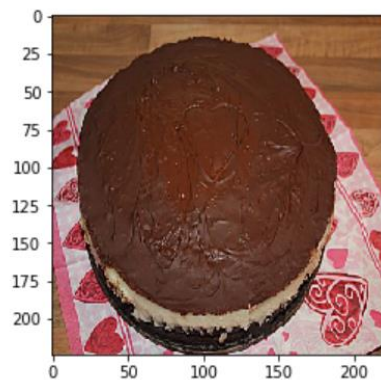


Рисунок 4.11 – Приклад зображення для побудови вектора

Тепер все, що потрібно зробити, це для кожного нового даного вхідного зображення пройти його через VGG-16, отримати вектор відбитків пальців і обчислити найближчих сусідів з приблизним пошуком найближчого сусіда. Бібліотека, яку будемо використовувати для цього `falconn`. `Falconn` — це бібліотека з алгоритмами для проблеми пошуку найближчого сусіда. Алгоритми

в falconn засновані на локально-чутливому хешуванні (LSH - Locality-Sensitive Hashing), який є популярним класом методів пошуку найближчого сусіда в високовимірних просторах. Метою falconn є забезпечення дуже ефективних і добре перевірених реалізацій структур даних на базі LSH.

В даний час falconn підтримує два типи LSH для подібності: гіперплан LSH і крос-політоп LSH. Обидва типи реалізуються з мульти-зондуванням LSH для того, щоб звести до мінімуму використання пам'яті. Крім того, falconn оптимізований як для щільних, так і для розріджених даних. Незважаючи на те, що він призначений для подібності козину, falconn часто може бути використаний для пошуку найближчих сусідів під евклідою або максимальним внутрішнім пошуком продукту.



Рисунок 4.12 – Зображення для тесування

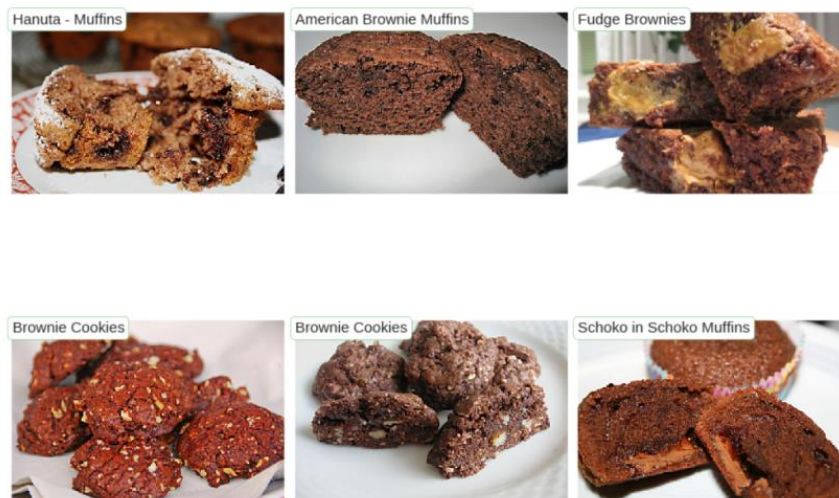


Рисунок 4.13 – Результат пошуку

Отримуємо пов'язані харчові зображення за зображенням запиту. Можемо створити сітку зображень для перегляду інтерпретації нейронної мережі. Наведене нижче зображення є лише невеликою частиною всього зображення. Можна побачити кулінарні страви, які мають подібні особливості ближче один до одного.

Висновки до розділу 4

Розпізнавання продуктів харчування кидає виклик поточним системам комп'ютерного зору, щоб вийти за рамки просто видимих. У порівнянні з природним розумінням зображення, прогнозування візуальних інгредієнтів вимагає високого рівня міркувань і попередніх знань (наприклад, що круасани, ймовірно, містять вершкове масло). Це створює додаткові проблеми, оскільки харчові компоненти мають високу верифікацію внутрішнього класу, важкі деформації відбуваються під час приготування їжі, а інгредієнти часто входять до приготованого блюда. Система є першим кроком до більш широких систем розуміння харчових продуктів, таких як оцінка калорій і створення рецептів.

Крім того, цей вид навчання може бути використаний для будь-якої проблеми, яка вимагає прогнозування довгоструктурованого тексту з зображення та прогнозованих ключових слів. Перша частина (прогнозування інгредієнтів) може бути застосована для вирішення більш широких проблем, таких як прогнозування зображення по набору.

Загальні висновки

У цій роботі запропоновано алгоритм визначення рецептів за зображеннями та додатковою описовою інформацією а саме рецептурним описом. Цей алгоритм представлений алгоритмічною композицією, що складається з CNN. CNN забезпечує розпізнавання зображень та порівняння з іншими зображеннями. Також використання рецептів дозволяє за ознаками групування подібності уточнювати позиціонування зображення та уникати помилкового визначення рецептів за зовнішньою схожістю зображення. Цей алгоритм показав на справжню ймовірність виявлення, що дорівнює 82% для певних груп зображень.

Експерименти показали, що реалізований алгоритм забезпечує високу ефективність. Запропонований алгоритм можна успішно оптимізувати, а потім застосувати до виявлення об'єктів.

Нейронні мережі зарекомендували себе як потужний алгоритм для класифікації зображень. Тому в цей час стали активно використовуватися та для інших задач машинного навчання.

Використання згорткових нейронної мережі для профілювання дозволить користувачам мережі Інтернет витратити менше часу на перегляд інформації і її обробку, а більше на її практичне застосування.

Представлені алгоритми визначення профілю зображення з використанням згорткової нейронної мережі, є абсолютно новими методами класифікації. Ефективність роботи алгоритму із семантичною репрезентацією набагато вище, чим при використанні векторної. Цей метод працює набагато ефективніше стандартного методу класифікації Bag of Words & TF IDF для звичайних нейронних мереж, тому що використовують у своїй основі згорткових нейронні мережі. Це досягнуто завдяки використанню переваг згорткових нейронних мереж у порівнянні з повнозв'язними нейронними мережами. Також

важливо, що в алгоритмі семантичної репрезентації ефективність вище в порівнянні з алгоритмом векторної репрезентації. Це пов'язане з використанням кластеризації і одержанню не деградуючого, при виконанні логічних операцій, вектора всього об'єкту.

Таким чином, аналізуючи дані, надається можливість спрогнозувати та запропонувати користувачеві найбільш необхідну для нього інформацію.

У результаті дослідження, при реалізації даних підходів стає можливим:

- підвищити ефективність пошуку та одержання інформації для користувача. Це стало можливим завдяки попередній обробці розробленими в цьому дослідженні алгоритмами та їх практичному застосуванню;

- підвищити точність розпізнавання рецептів завдяки використанню не тільки зображень а також рецептів, що надає мережі додаткові особливості за якими можливо більш точно позиціонувати зображення;

- розробити систему розпізнавання інгредієнтів блюда базуючись суто на зображенні;

- використання додаткової інформативності а саме, описової інформації щодо рецепту блюда дає змогу запропонувати користувачеві ряд схожих рецептур;

- розширити інформацію щодо зображення шляхом використання додаткової інформації з рецепту приготування блюда.

Таким чином, сфера можливого застосування результатів наукового дослідження досить широка, що показує високу цінність проробленої роботи.

Перелік посилань

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., Drozdal, 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).

Додатки

```
def write_recipe_details(data):
    dpath = DATA_FOLDER + DFILE_NAME
    with open(dpath, 'a', newline='') as f:
        writer = csv.writer(f)
        try:
            writer.writerow((data['link'],
                             data['ingredients'],
                             data['zubereitung'],
                             data['tags'],
                             data['gedruckt:'],
                             data['n_pics']
                             #data['reviews'],
                             #data['gespeichert:'],
                             #data['Freischaltung:'],
                             #data['author_registration_date'],
                             #data['author_reviews']
                             ))
        except:
            writer.writerow('')
```

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 розпізнавання дозволяє зробити крок до систем розуміння продуктів харчування, таких як оцінка калорій та створення рецептів. Система розпізнавання 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 вивчалось 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 ранжированих списках.

Формула обчислення коефіцієнта рангової кореляції Кендалла може бути виражена як:

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

де $P(p)$ - кількість збігів;

$P(q)$ - кількість інверсій;

N - обсяг вибірки.

У спрощеному вигляді формулу коефіцієнта кореляції Кендалла можна записати як:

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

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

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

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 .

$$\begin{aligned} K_X &= \sum_i |X_i| (|X_i| - 1) \\ K_Y &= \sum_j |Y_j| (|Y_j| - 1) \end{aligned} \quad (4)$$

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 $N \times 4096$ до матриці $N \times 512$. 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. В 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:

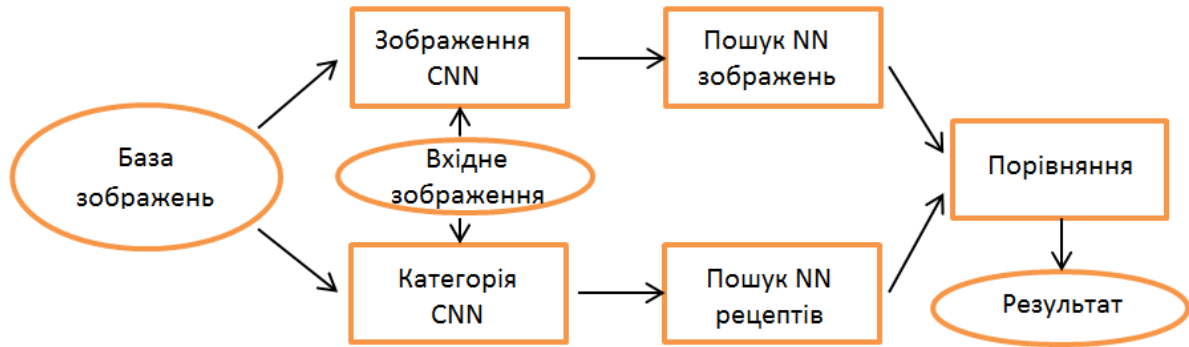


Рис. 1 Схеми функціонування системи розпізнавання зображень

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 проходить 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 відповідає 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 азійський 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 $n \times 4096$, де n - кількість вхідних зображень.

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

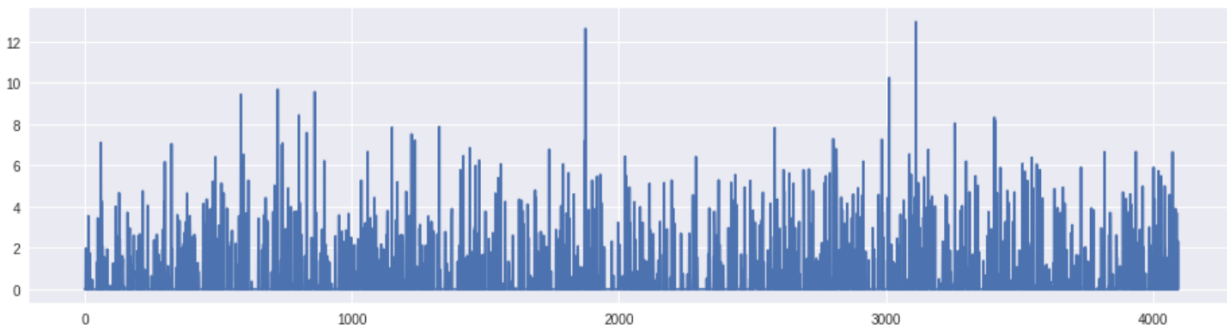


Рис. 2. Вектор 4096 обчислений від зображення тістечка

Тепер все, що потрібно зробити - це 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., Drozdal, 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

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

МАГІСТЕРСЬКА РОБОТА

Використання штучного інтелекту
для оцінювання розпізнавання зображень

Розробив ст. гр. КНм-19-1:
Гірний М.Ю.

Хмельницький - 2020

В магістерській роботі розроблений і реалізований метод отримання інформації для розпізнавання рецептів та інгредієнтів за зображенням блюд та їх рецептурою. Розроблювальна система пропонує користувачам блюда основі виявлених схожих, які виражені в текстовій формі та у вигляді зображень. Така система дозволяє за зображення згенерувати рецепт та скласти список інгредієнтів базуючись на подібності зображення та схожості найближчих рецептів.

Метою дослідження є розробка методу отримання інформації для систем автоматичного формування рецептів базуючись на зображенні блюда.

Для досягнення зазначеної мети поставлені наступні **задачі**:

- показати, що використання візуальної інформації та додаткової текстової інформації дозволяє створити систему розпізнавання рецептів а, відтак і інгредієнтів, базуючись на зображенні блюда;
- провести дослідження ознак і класифікаторів, застосовуваних в машинному навчанні в задачах розпізнання складових елементів та інтерпретації зображень;
- провести порівняння застосовності методів машинного навчання, до задачі отримання інформації про зображення базуючись на подібності зображення та подібності описової текстової інформації, в даному випадку рецептів.

Задачі:

- попередня обробка колекції рецептів та зображень страв;
- побудова словників рецептів;
- побудова шаблонів рецептурних форм та зображень та розміщення їх за групами;
- побудова кластерів подібних страв за зображенням та описовими формами - інгредієнтами;
- розробка ймовірнісних моделей групування зображень;
- виділення тих параметрів, для отримання яких може бути застосоване машинне навчання (на основі розподілу значень параметрів в розміченому підкорпусі);
- вибір моделей, виділення ознак і застосування методів машинного навчання для отримання інформації про відповідні параметри;
- тестування методів на основі правил і з використанням машинного навчання та вибір оптимального методу для кожного з параметрів;

Об'єктом дослідження є методи отримання інформації із зображень та описової текстової інформації.

Предметом дослідження є зображення страв, а також описи рецептів та відповідні інгредієнти, що є додатковою описовою інформацією до зображень.

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.

Практична значимість дослідження полягає в тому, що описані методи та отримані результати можуть бути використані при розробці систем автоматичного генерування рецептів на базі зображень.

В результаті проведеного дослідження показано, що за допомогою даних, отриманих на основі аналізу зображень та рецептурного опису, може бути підвищена ефективність роботи системи у розпізнанні зображень базуючись також на додатковому описі тобто текстовою інформацією до зображення.

Що стосується експериментів з машинним навчанням, в роботі показано експериментальні дослідження запропонованого методу навчання базуючись на зображенні та текстовому описі.

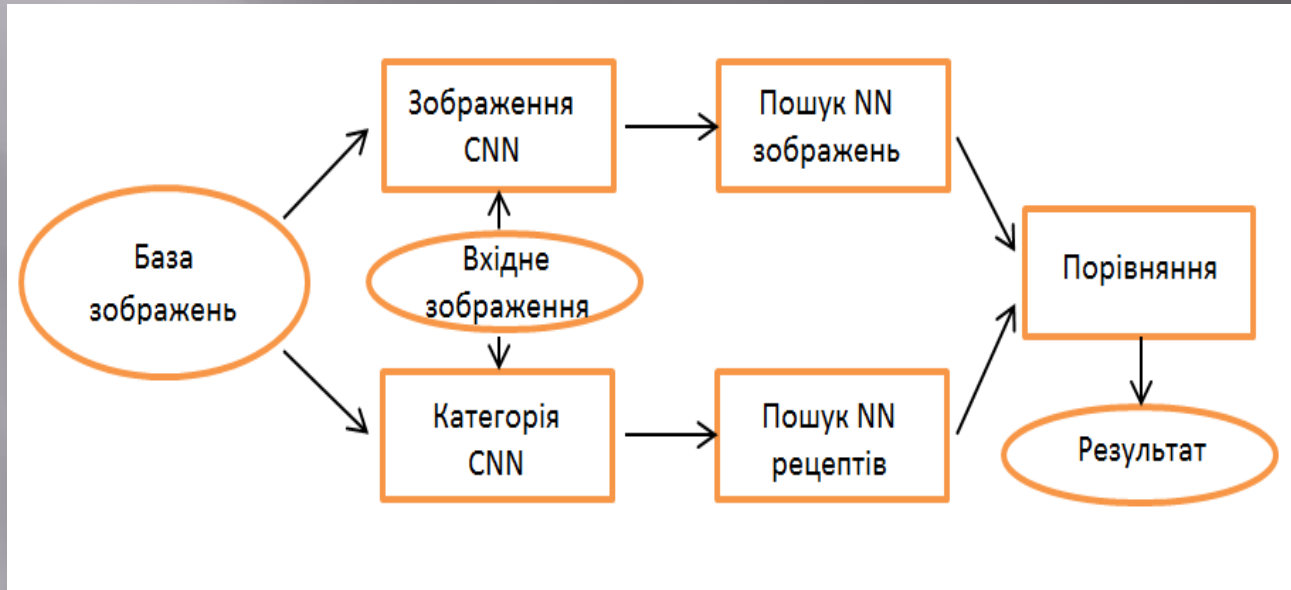


Схема функціонування системи розпізнавання зображень

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 Діріхте (LDA)

- Матрична факторизація

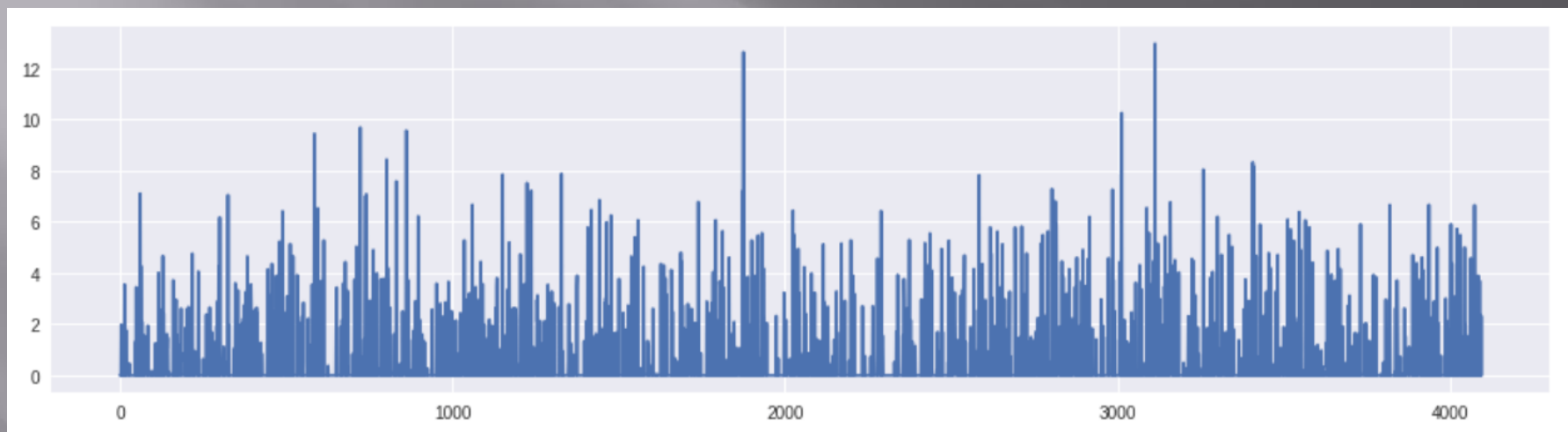
4 Добування особливостей

- k-найближчі сусіди

- Візуалізація t-SNE

5 Трансферне навчання: Навчання заздалегідь підготовленої CNN (конволюційна нейронна мережа)

Вектор обчислений від зображення тістечка 4096



Висновки

У цій роботі запропоновано алгоритм визначення рецептів за зображеннями та додатковою описовою інформацією а саме рецептурним описом. Цей алгоритм представлений алгоритмічною композицією, що складається з CNN. CNN забезпечує розпізнавання зображень та порівняння з іншими зображеннями. Також використання рецептів дозволяє за ознаками групування подібності уточнювати позиціонування зображення та уникати помилкового визначення рецептів за зовнішньою схожістю зображення. Цей алгоритм показав на справжню ймовірність виявлення, що дорівнює 98,1% для певних груп зображень.

Експерименти показали, що реалізований алгоритм забезпечує високу ефективність. Запропонований алгоритм можна успішно оптимізувати, а потім застосувати до виявлення об'єктів.

Дякую за увагу

Anti-Plagiarism v-15.257

Максимальне співпадіння з одним документом 2.0%

Словники перевірки: en_US, ru_RU, ua_UA. **Помилоч в документах: 6%**

ID: 81681 Назва: Використання штучного інтелекту для оцінювання розпізнавання зображень Додано в БД: 2020-11-30 Автора: Гірний Микола Юрійович Керівники: Манзюк Е.А. Консультанти: Опоненти:	Документ		Сумарний збіг по Базі Даних	
	Символи	Лексеми	Символи	Лексеми
	67272	554	1903 (3%)	14 (3%)

Джерело плагіату

ID	Опис	Наявність плагіату в документі	
		Символи	Лексеми

РІШЕННЯ КАФЕДРИ КОМП'ЮТЕРНИХ НАУК ТА ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ ПРО ДОПУСК КВАЛІФІКАЦІЙНОЇ РОБОТИ ДО ЗАХИСТУ

Підтверджуємо ознайомлення з результатом звіту подібності щодо роботи, генерованого системою виявлення текстових збігів/ідентичності/схожості:

Назва: **Використання штучного інтелекту для оцінювання розпізнавання зображень**

Автор: **Гірний М. Ю.**

Спеціальність: **122 Комп'ютерні науки**

Науковий керівник: **к.т.н. доцент Манзюк Е.А.**

Після аналізу звіту подібності зроблено такий висновок:

№	Висновок	Позначка про відповідність
1	Запозичення, виявлені в роботі, є законними і не є плагіатом (далі – зазначаються підстави віднесення запозичень до правомірних). Робота приймається до захисту.	відповідає
2	Виявлені запозичення не є плагіатом, розміщені в розділах, які не описують безпосередньо авторське дослідження, але кількість цитат перевищує обсяг, виправданий поставленою метою роботи (далі – зазначаються детальні та аргументовані підстави віднесення запозичень до правомірних). Робота приймається до захисту, але має бути відкоригована. Відкоригований варіант має бути поданий на кафедру за 2 дні до захисту, разом із заявою щодо самостійності виконання письмової роботи та ідентичності друкованої та електронної версії роботи	-
3	Робота містить навмисні текстові спотворення, передбачувані спроби укриття запозичень або інші прояви академічного плагіату. Робота містить фабрикацію або фальсифікацію даних. Робота не допускається до захисту.	-
4	Інше:	-

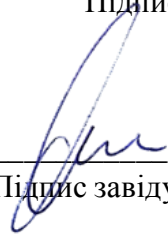
Підтвердження: Виявленні запозичення не є плагіатом так як не описують безпосередньо авторське дослідження і є широко вживаними поняттями предметної області і складають 5.9%.

01.11.2020

Дата



Підпис керівника



Підпис завідувача кафедри

ВІДГУК ОПОНЕНТА
на дипломну роботу магістра

Магістра гр. КНм-19-1 Гірного Миколи Юрійовича

На тему Використання штучного інтелекту для оцінювання розпізнавання зображень

1. Актуальність і значення теми

В магістерській роботі розроблений і реалізований метод отримання інформації для розпізнавання рецептів та інгредієнтів за зображенням блюд та їх рецептурою. Розроблювальна система пропонує користувачам страви на основі виявлених схожих, які виражені в текстовій формі та у вигляді зображень.

2. Оцінка якості та достовірності проведених досліджень

Достовірність результатів забезпечується проведенням всебічного оцінювання та порівняння ефективності різних методів, в тому числі за допомогою статистичних тестів.

3. Оцінка запропонованих заходів та пропозицій, практичної цінності та ефективності

Практична значимість дослідження полягає в тому, що описані методи та отримані результати можуть бути використані при розробці систем автоматичного генерування рецептів на базі зображень.

4. Загальні висновки та оцінка

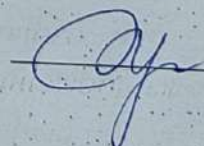
В результаті проведеного дослідження показано, що за допомогою даних, отриманих на основі аналізу зображень та рецептурного опису, може бути підвищена ефективність роботи системи у розпізнанні зображень базуючись також на додатковому описі тобто текстовою інформацією до зображення.

Що стосується експериментів з машинним навчанням, в роботі показано експериментальні дослідження запропонованого методу навчання базуючись на зображенні та текстовому описі. В роботі існують певні недоліки, які полягають у відсутності семантичних зв'язків між словами, що знижує ефективність системи.

Дипломна робота заслуговує на оцінку визначено

Рецензент (прізвище, ім'я, по батькові, посада, місце роботи) Медведчук М.Р.
к.т.н., доцент кафедри автоматизованого машинобудування та агрегації енергії

“3” 11 2020р.

 (підпис)