DSC180B Capstone Final Project: PicToPlate

From Ingredients to Recipes with Retrieval Augmented Generation

Megan Huynh
mlhuynh@ucsd.edu

Christine Xu
jix028@ucsd.edu

Colin Jemmott
cjemmott@ucsd.edu

Abstract

The process of deciding what and how to cook, especially when trying to optimize the use of available ingredients, can be confusing and time-consuming. Our product will address this practical need in everyday life by leveraging the power of object detection and large language models to simplify and enhance the cooking experience for home cooks. Utilizing GPT-4 Vision, our product will allow users to input an image of groceries, create a list of available ingredients, and then search a database of recipes that use those ingredients. Additionally, and AI chatbot will guide users through each step of the process, such as allowing user input for additional ingredients and guiding them through the steps of the recipe. This innovative approach combines computer vision, search, and large language models to enhance the user experience in the kitchen.

Code: https://github.com/ChristineXu0924/recipe_retrieval

1 Introduction

1.1 Motivation

For those looking to minimize time spent searching for recipes, Google is simply not sufficient. It fails to return succinct results, and does not adequately address search personalization, such as filtering out recipes that include a certain ingredient. In our project, we will utilize object detection models and large language models to enhance recipe search and add assistance throughout the cooking process.

Samsung, a major fridge retailer, has been developing fridges that contain internal cameras(1). These cameras, paired with various artificial intelligence (AI) image classifiers, will recognize what items are stored, create an internal inventory, then create personalized recipe recommendations. Fittingly named 'smart fridges' also have the ability to learn about the user, such as how quickly they use an ingredient and are also able to create shopping lists according to the internal inventory. These features are all powered by Samsung Food "the ultimate cooking app for recipe saving, meal planning, grocery shopping, and recipe sharing". Launched in 2023, the technology has been in development for a few years and has been integrated into their fridges.

As novel as these smart fridges may be, the average consumer may not want to spend thousands of dollars on a new fridge. Creating a free application that will simplify and optimize recipe search will impact a wider range of people.

1.2 Goals

We hope to create an interactive application where a user can input pictures of their groceries to communicate with our AI chat bot for guidance in generating cooking recipes ideas and executing the dish. Our project will perform the following steps in data analysis and result production:

1. Object detection from a user inputted image
2. Search using the list of ingredients returned from the previous step in a database of recipes
3. Display possible recipes to cook, involving recipe summarization and further interaction

1.3 Literature Review and Discussion of Prior Work

This problem was examined very recently in the paper "Food Recipe Recommendation Based on Ingredients Detection Using Deep Learning"(2) published in 2022. In this paper, they divided the problem into two parts: object recognition and recommendation, both rooted in various open-source machine learning and deep learning approaches. The researchers first had to begin with object recognition in order to classify ingredients presented in an

image. This was achieved by training a convolutional neural network (CNN) trained on the ResNet50 dataset, a collection of more than 1 million images. The CNN was then able to classify images into different categories of food, such as "beef meat, cheese, onion, pasta," etc. Then, the next step was to make recipe recommendations based on these ingredient classes. They created a 2-D matrix mapping the one-to-one relationships, between different ingredients to identify which recipes can be cooked with which ingredients. Linear search was then performed using these relationship and their model achieved an accuracy of 94%.

CNNs seem to be a common approach in object recogntion; however, the recent release of GPT-4 Vision(3) seems to be a strong contender to accurately perform this task. We would like to experiment with this influential cutting-edge technology.

1.4 Dataset Description

We merged the Kaggle "Food.com Recipe and Interaction" dataset with the RecipesNLG dataset. The "Food.com Recipe and Interaction" dataset includes a recipes table which has recipe name, tags, nutrition, n_steps, step details, description; and a user_interaction table which provides date_of_interaction, rating, and review. The dataset was originally uploaded by Food.com. The RecipesNLG dataset, from Hugging Face, also contains many of the Food.com recipes; however it also contains other websites and most importantly, contains measurements for ingredients. Since measurements are vital for cooking, we decided to merge the two datasets to add the measurements. After the merge, we only used data from Food.com

- `name`: name of the recipe
- `id`: unique identifier for each recipe
- `minutes`: time in minutes to cook the recipe
- `tags`: recipe descriptors, such as type of cuisine, breakfast/lunch/dinner/snack, quick, etc.
- `nutrition`: number of calories, total fat in percent daily value (PDV), sugar (PDV), sodium (PDV), protein (PDV), and saturated fat
- `ingredients`: a list of ingredients needed for making this recipe. Including the exact measurement.

2 Methods

2.1 Pre-Processing

Aside from merging the datasets as previously mentioned, there was little data-preprocessing that we implemented. We used a named entity recognition (NER) model to extract the key ingredients from each recipe, removing the redundant descriptions and measurements. For example, an ingredient listed as "1 (10 ounce) can prepared pizza crust" will be parsed into "pizza crust". Removing redundancies improves the performance of our keyword search

(TF-IDF) algorithm.

2.2 Object Detection Model

The model we used to determine individual ingredients in an image is the GPT-4 Vision model. We also tried two open source models, ResNet-50 and YOLO, however, GPT-4 performed significantly better. GPT-4 is able to accurately identify almost every ingredient in the images we have tested. However, a limitation of the model is that it is unable to detect the quantity of ingredients present other than counting.

2.3 Search

We tested 3 different options for our search model: semantic search, TF-IDF, and using `str.contains()` and evaluated them qualitatively. We compared how each search algorithm performed by comparing the ingredients in the recipe returned against our list of input ingredients. From our tests, we found that semantic search made too many assumptions about our inputs, and returned recipes used our ingredients, but also required many other ingredients that were not included in the input. Using `str.contains()` seemed to have to opposite effect; the recipes returned matched our input, but were extremely simple, often being beverages that only required 2-3 ingredients.

TF-IDF was in the middle of the spectrum: the recipes matched our input list, but made some inappropriate ingredient substitutions. For instance, cream cheese in a cheese cake recipe should not be replaced with cheddar. In order to correct for this effect, we trained a Word2Vec(4) embedding that captures distributional similarities between words. That is, words that often appear in similar contexts are mapped to vectors separated by a shorter Euclidean distance (the L2 norm). In our case, the words here are recipe ingredients. The following is one example: mozzarella cheese, which is most commonly used in pizza, is considered similar with other pizza-related cheese and ingredients.

```
model = Word2Vec.load("models/model_cbow_2.bin")
model.wv.most_similar(u'mozzarella cheese')
```

✓ 0.4s

```
[('skim mozzarella cheese', 0.8097931742668152),
 ('lowfat mozzarella cheese', 0.7782601714134216),
 ('cheese blend', 0.7505210041999817),
 ('pre mozzarella cheese', 0.7094582319259644),
 ('reduced mozzarella cheese', 0.7030929923057556),
 ('pizza cheese', 0.6895893812179565),
 ('provolone cheese', 0.6858465671539307),
 ('pepperoni', 0.6494641900062561),
 ('mozzarella', 0.6118200421333313),
 ('old cheddar cheese', 0.5831028819084167)]
```

Figure 1: Example of terms similar to “mozzarella cheese”

2.4 Retrieval-Augmented Generation (RAG) with LangChain

Our final part invovles using LLMs to support conversations with user. LangChain(5) is a framework for deploying LLMs for applications, and in our case, the context-aware generation. The top related recipes retrieved from our search, as well as their descriptions, will be injected into the LLM (GPT-3.5-turbo) as reference when answering questions and fulfilling users' requirements. The store and load of chat history is one of the biggest challenges in implementing this feature.

3 Search Evaluation

Search is inherently hard to evaluate due to two conflicting factors: usefulness and accuracy. To first account for accuracy, we calculated the cosine similarity scores between our ingredient query list and the ingredients utilized in the recipe. However we also felt that it was important to qualitatively examine the results of our search based on if the returned recipe could be feasibly cooked with the ingredients in the query. We will test a number of different ingredient queries, count how many matches there are between the recipe's ingredients and the query's ingredients, and calculate a "feasibility" score. These matches are also subjective on the criteria of "with the ingredients in the query, will the final product closely resemble the target recipe?" This subjective evaluation allows for ingredient substitution and takes into account common spices or condiments found in a typical kitchen, such as as salt, pepper, sugar, water, and others.

4 Application Deployment

We used Streamlit to host our application due to its features that align with our product goals. Users can upload their own images, or use camera input within the Streamlit site. Streamlit was primarily used as a front-end service for user interaction/experience. Additionally, this is where we hosted our GPT-4 chatbot.

5 Results

5.1 Search

We created 4 sample ingredient queries to evaluate our search. Consistently, the cosine similarities scores were considerably lower than our feasibility score.

The ingredient samples are as follows:

1. pasta, olive oil, tomato sauce, cheese, mushrooms, zucchini, chicken, onion, garlic
2. steak, oil, onion, egg, garlic, honey, rice, butter, mushroom

Table 1: Comparison Between Cosine Similarity and Feasibility Score for the Top 10 Search Results

Sample	Average Cosine Similarity	Average Feasibility Score
Ingredient Sample 1	0.5729077869	0.8258874459
Ingredient Sample 2	0.5157574193	0.875
Ingredient Sample 3	0.7384241854	0.6964285714
Ingredient Sample 4	0.4771470249	0.7428571429

3. coffee, milk, sugar, whipped cream, chocolate, vanilla
4. romaine lettuce, spinach, kale, cheese, tomato, balsamic vinegar, egg, salad dressing, olives, croutons

6 Discussion

A key challenge we faced was the integration of object detection technology to identify ingredients from user-uploaded images. We experimented with different models, including GPT-4 Vision, ResNet-50, and YOLO, to accurately detect individual ingredients. While GPT-4 Vision performed exceptionally well in ingredient recognition, we also faced challenges in quantifying ingredient quantities, which is crucial for recipe preparation.

Another aspect of the project that we struggled with was the balance between accuracy and usefulness in recipe search. While our search algorithm aimed to return recipes that closely matched the user's input ingredients, we also had to consider the adaptable nature of cooking. Oftentimes ingredients in a recipe can be closely substituted with similar ingredients or some ingredients may not be essential to cooking the intended product. This contradiction required a subjective evaluation that took into account common substitutions and pantry staples, ensuring that the recommended recipes were practical and achievable for the user. While the cosine similarity scores were quite strict, we often found that cooking the recipe was still in fact feasible.

Furthermore, our application deployment on Streamlit allowed for user interaction and integration of a chat bot powered by GPT-4. While there was real-time guidance and assistance throughout the cooking process, we failed to optimize our code to be hosted on a cloud service. Our loading times ended up being quite slow, which negatively impacted the user experience.

7 Conclusion

Our project, PicToPlate, addresses the practical need for simplifying and optimizing the recipe search experience for home cooks. By leveraging cutting-edge technologies such as GPT-4 Vision, fine-tuned search models, and retrieval augmented generation, we have

developed an innovative solution that allows users to input images of their groceries, receive personalized recipe recommendations, and obtain real-time guidance through an AI chatbot. We are happy with what we accomplished in 10 short weeks.

Throughout the development process, we encountered challenges such as accurately detecting ingredient quantities, balancing search accuracy with usefulness, and considering UI/UX when building an application. Additionally, due to how multifaceted and ambitious our project was, it was difficult to decide how to allocate time and effort to the different parts of the whole task.

In the future, there are many improvements to be made to our product. We could try developing our own object detection model specifically suited to recognizing food items. Additionally, we would like to continue fine tuning our search to achieve higher cosine similarity scores. Lastly, there are many UI/UX elements to be improved upon, such as decreasing the amount of loading/wait time on the application.

References

[1] Samsung Newsroom (2020). *New Food AI Looks Inside Your Fridge To Help You Find The Perfect Things To Cook With What You ALREADY Have*.

[2] Md. Shafaat Jamil Rokon and Md Kishor Morol and Ishra Binte Hasan and A. M. Saif and Rafid Hussain Khan (2022). *Food Recipe Recommendation Based on Ingredients Detection Using Deep Learning*.

[3] OpenAI GPT-4 Vision. `https://platform.openai.com/docs/guides/vision`.

[4] TensorFlow word2vec. `https://www.tensorflow.org/text/tutorials/word2vec`.

[5] LangChain retrieval. `https://www.langchain.com/retrieval`.