A Web-Based Personalized Recipe Recommendation System

Liam Li[*]
zl127@illinois.edu
University of Illinois
Urbana-Champaign
Champaign, Illinois, USA

Lin Guo
linguo4@illinois.edu
University of Illinois
Urbana-Champaign
Champaign, Illinois, USA

Yifan Zhong
yzhong32@illinois.edu
University of Illinois
Urbana-Champaign
Champaign, Illinois, USA

Lingge Wu
linggew2@illinois.edu
University of Illinois
Urbana-Champaign
Champaign, Illinois, USA

Zixiong Luo
zl130@illinois.edu
University of Illinois
Urbana-Champaign
Champaign, Illinois, USA

ABSTRACT

In the modern quest for tailored culinary experiences, we develop an innovative web-based recipe recommendation system designed to deliver personalized recipe suggestions to users based on their unique dietary preferences, restrictions, and available ingredients. This cutting-edge platform stands to redefine home cooking by offering a user-friendly interface for inputting food preferences in natural language, followed by an intelligent feedback-informed algorithm that curates recipes to match individual needs.

CCS CONCEPTS

• **Information systems** → Recommender systems; Business intelligence; Language models; Personalization; Collaborative search; Search interfaces; • **Computing methodologies** → Information extraction.

KEYWORDS

Recommendation system, Information retrieval and ranking, Natural language processing

1 INTRODUCTION AND MOTIVATION

The proposed web-based recipe recommendation system is a novel solution designed to enhance personal culinary experiences by addressing key pain points in current recipe discovery processes. Here's why this tool is significant and necessary:

- **Personalization:** It directly caters to individual dietary preferences and restrictions, offering a tailored recipe selection that current platforms lack, thus solving the challenge of sifting through irrelevant options.
- **Efficiency and Waste Reduction:** By suggesting recipes based on available ingredients, it promotes efficient use of resources, potentially reducing food waste and aiding in environmental sustainability.
- **Health and Dietary Support:** The system supports health and wellness goals by ensuring that all recommended recipes align with users' dietary needs, contributing to better health outcomes.

[*]Coordinator

In summary, this tool not only enhances the meal planning process for individuals with specific dietary needs but also contributes to broader societal goals of sustainability and wellness.

2 USERS

The target users are individuals seeking culinary inspiration that matches their specific dietary needs, available cooking ingredients, or desire to explore new recipes. This includes, but is not limited to, home cooks, individuals with dietary restrictions, and those looking to diversify their meals.

3 FUNCTION

Our group implements four functions for this recommendation system:

(1) Select a random recipe from the drop-down list, and generate recommended recipes based on our recommendation metric.
(2) Enter several ingredients that you want to cook with and several ingredients that you don't want to cook with, and generate recommended recipes based on our recommendation metric.
(3) Enter natural language that describes the ingredients that you want to cook with, and generate recommended recipes based our recommendation metric.
(4) Search and rate a recipe.

Among these four functions, the first three are the recommendation functions, the last one is related to our recommendation metric, see more details in the Implementation section.

For recommendation functions, the first one is useful when the user knows a recipe and wants more recipes with similar ingredients and high ratings. The second one is useful when the user knows what ingredients he/she wants/doesn't want to cook with and wants more recipes based on these conditions. The third one is useful when the user doesn't know the exact name of ingredients, then he/she can enter natural language to describe them.

4 IMPLEMENTATION

4.1 Data

The original data is "Food Ingredients and Recipes Dataset with Images" from Kaggle [1]. It has 13582 rows and 5 columns. Our group investigate and utilize 4 columns which are "Title", "Ingredients", "Instructions", and "ImageName".

4.2 Data Preprocessing

To better implement the major functions of our recommendation system, our group preprocess the "Ingredients" column to ensure that it only contains relevant words. The order of the preprocessing is as follows:

(1) **Extract Words (Tokenization)**: We use regular expression to extract lowercase words by removing punctuation.
(2) **Lemmatization**: We use spaCy to lemmatize the extracted words.
(3) **Remove stop words**: We use NLTK to remove common English stop words.
(4) **Remove Numbers**: We use regular expression to remove integers, floats, and unicode fractions.
(5) **Remove Cooking Metrics**: We use regular expression to remove common cooking metrics [2].
(6) **Remove Common Household Items**: After the above five steps, we find out the top 50 frequent words among all the ingredients and consider them as common household items. Then we use regular expression to remove these common household items.

By doing such preprocessing, it is more accurate to calculate the similarity of ingredients among different recipes.

4.3 Model

Our recommendation system belongs to content-based filtering, it utilizes TF-IDF model and Cosine Similarity model to implement the functions:

(1) **TF-IDF**: First, we use TF-IDF to vectorize the preprocessed "Ingredients" column.
(2) **Cosine Similarity**: Second, we use cosine similarity to calculate the similarity between different vectorized values.

By using TF-IDF and Cosine Similarity, we can easily implement our major functions. During the evaluation process, we try different feature extraction techniques besides TF-IDF, including BM25, Word2Vec, and Bert, see more details in the Evaluation section.

4.4 Recommendation Function

As mentioned in the Function section, our group implements four functions for this recommendation system and the first three of them are real recommendation functions, their implementation details are as follows:

- **Function 1**: The user selects a random recipe from the drop-down list. Then the system uses TF-IDF to vectorize the preprocessed "Ingredients" column of the selected recipe and all other recipes in the dataset. Next the system calculates the cosine similarity and ranks the recipes based on our recommendation metric. Finally the system generates top 5 recipes as the recommended recipes.

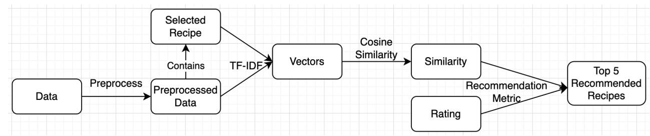

Figure 1: Structure of Function 1

- **Function 2**: The user enters several ingredients that he/she wants to cook with and several ingredients that he/she doesn't want to cook with. The system first filters out all the recipes that contain ingredients that the user doesn't want to cook with. Then the system uses TF-IDF to vectorize the input ingredients that the user wants to cook with and the preprocessed "Ingredients" column of the remaining recipes in the dataset. Next the system calculates the cosine similarity and ranks the recipes based on our recommendation metric. Finally the system generates top 5 recipes as the recommended recipes.

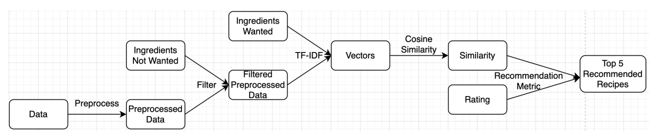

Figure 2: Structure of Function 2

- **Function 3**: The user enters natural language that describes the ingredients that he/she wants to cook with. The system first uses a keyword extraction algorithm from Hugging Face [3] to extract relevant ingredients. Then the system uses TF-IDF to vectorize the extracted relevant ingredients and the preprocessed "Ingredients" column of the recipes in the dataset. Next the system calculates the cosine similarity and ranks the recipes based on our recommendation metric. Finally the system generates top 5 recipes as the recommended recipes.

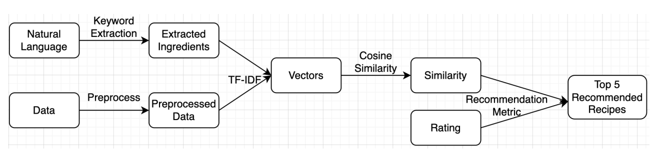

Figure 3: Structure of Function 3

[1] https://www.kaggle.com/datasets/pes12017000148/food-ingredients-and-recipe-dataset-with-images/data
[2] https://en.wikibooks.org/wiki/Cookbook:Units_of_measurement
[3] https://huggingface.co/ilsilfverskiold/tech-keywords-extractor

4.5 Recommendation Metric

For all the recommendation functions that we implement, they generate the recommended recipes based on our recommendation metric. For most of the recommendation systems, the recommendation metric only considers the similarity. The higher the similarity, the higher the ranking, and thus the higher the possibility to recommend. However, for our recommendation system, we design a new recommendation metric that combines both similarity and rating, the formula is as follows:

$$Score = \alpha \cdot Similarity + \beta \cdot Rating \tag{1}$$

$$\alpha + \beta = 1 \tag{2}$$

Here Score is the weighted sum of Similarity and Rating, the higher the Score, the higher the possibility to recommend! α and β are weights that reflect the relative importance of Similarity and Rating, their values depend on the user's preference. For example, if the user cares more about Similarity, then the value of α can be 0.8, and the value of β should be 0.2.

5 EVALUATION

Since the evaluation of a recommendation system is very difficult because the accuracy of the recommended results depend on personal preference, instead of using surveys to manually test the system, our group implements a GPT-4 based Recipe Recommendation System with the same dataset as the baseline for evaluation. This GPT-4 based recommendation system operates as a standard, we will provide the same input value to our original system and this LLM based system and compare the results.

Moreover, since our original recommendation system uses TF-IDF to extract features, our group also tries other feature extraction techniques, including BM25, Word2Vec, and Bert.

Since our group implements three recommendation functions and the second one is similar to the third one, we only tested the accuracy of the first function and the second function among TF-IDF, BM25, Word2Vec, Bert, and GPT-4, the evaluation result is as follows:

	TF-IDF	BM25	Word2Vec	Bert	GPT-4
Function 1	0.93	0.91	0.95	0.97	1
Function 2	0.34	0.35	0.47	0.53	1

Table 1: Evaluation Results

For Function 1, Bert exhibits the highest accuracy at 0.97, closely followed by Word2Vec at 0.95, and TF-IDF at 0.93. BM25 shows a lower performance at 0.91. This suggests that Bert's context-aware embeddings are particularly effective for capturing the nuances in recipe data, which likely contributes to more precise recipe matching and recommendation.

For Function 2, Bert leads with 0.53 accuracy, indicating its superior capability to discern and utilize ingredient-related context within recipes. Word2Vec also performs reasonably well at 0.47, suggesting its effectiveness in capturing semantic similarities. TF-IDF and BM25 show lesser effectiveness for this specific filtering criterion, at 0.34 and 0.35 respectively. This may be due to their limitations in capturing the depth of semantic relationships between ingredients.

Bert's superior performance across both function 1 (random selection) and function 2 (ingredient-based filtering) highlights its effectiveness in understanding complex relationships and context within recipes, making it an ideal choice for content-based recommendation systems in culinary domains. While BM25 lags in this application, it might still be useful in scenarios where exact term matching is more critical.

6 CONTRIBUTION

- Liam Li: Collected and Preprocessed the data; Implemented the recommendation functions using TF-IDF and cosine similarity; Designed the recommendation metric.
- Lin Guo: Completed the project proposal and project presentation.
- Yifan Zhong: Implemented a GPT-4 based Recipe Recommendation System as the baseline for evaluation.
- Lingge Wu: Designed and developed the front end for this system.
- Zixiong Luo: Utilized different feature extraction techniques, including BM25, Word2Vec, and Bert to implement different Recipe Recommendation Systems as comparision groups for evaluation.

7 USER MANUAL

This section provides a detailed guide on how to effectively utilize the functions of our Recipe Recommendation System. Our Recipe Recommendation System has been published on GitHub [4].

7.1 Recipe Selection from Dropdown Menu

Begin by exploring our comprehensive dropdown menu, which lists all the recipes available in our database. Here's how to access and use this feature:

(1) Open the dropdown menu on the system's main interface.
(2) Browse through the list and select any recipe that interests you.
(3) Upon selection, the interface will not only display the selected recipe but will also recommend four additional similar dishes, enhancing your culinary exploration.

Note: The similar dishes are selected based on their ingredients and popularity, ensuring a variety of choices that adhere to similar taste profiles.

7.2 Custom Ingredient Filtering

Tailor your recipe search to fit your exact culinary needs with our custom ingredient filtering option. Follow these steps to filter recipes:

(1) Enter the ingredients you have at home or wish to use in the 'Preferred Ingredients' field.
(2) Specify any ingredients you dislike or need to avoid in the 'Disliked Ingredients' field.

[4] https://github.com/liamli99/CS510Project

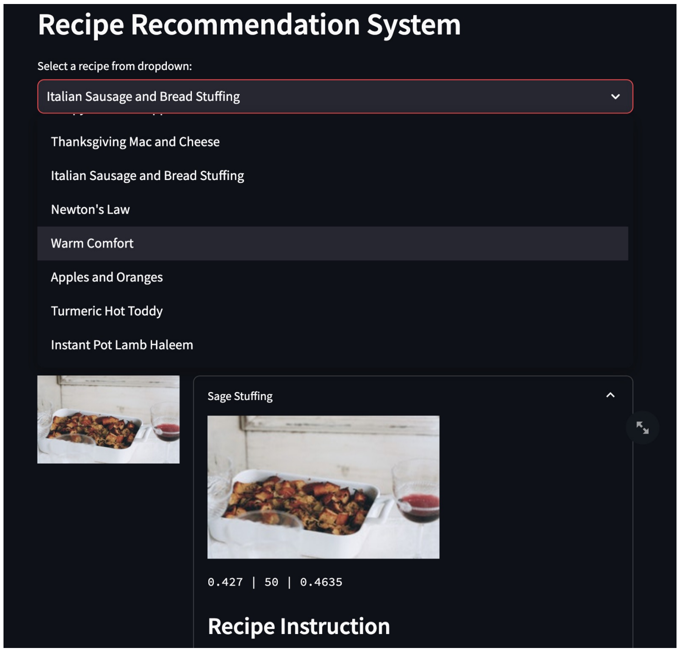

Figure 4: Dropdown menu with recipes selection.

(3) Click the 'Show Recommendation' button. The system will then display a list of recipes that match your preferences, excluding any recipes that contain your disliked ingredients.

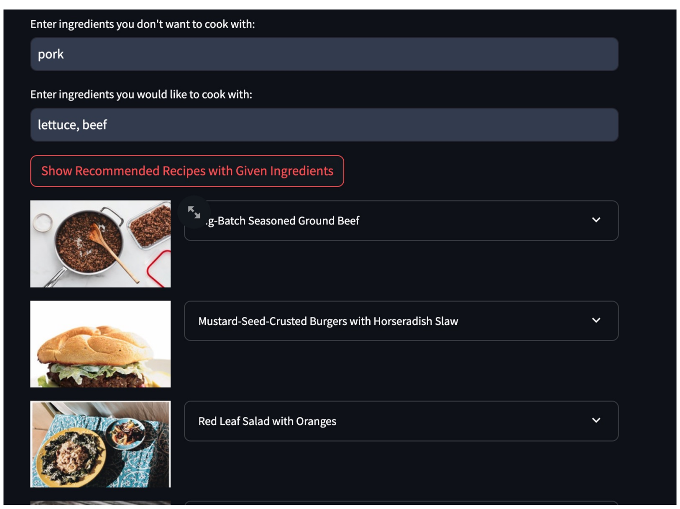

Figure 5: Interface for filtering ingredients in the recipe search.

This feature is particularly useful for avoiding allergens, adhering to dietary restrictions, or simply refining your search to match your current pantry.

7.3 Natural Language Query

For an even more intuitive search experience, use the natural language query feature. This allows you to input your ingredients or recipe requests in a conversational manner:

(1) Navigate to the query box on our interface.
(2) Input a natural language query such as, "What can I make with chicken and rice?"
(3) The system will analyze your input and instantly provide a list of the top five recipes that use those ingredients.

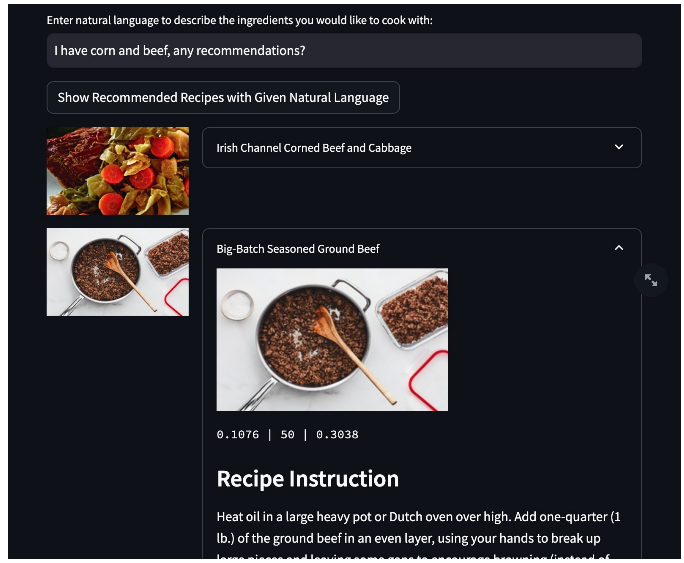

Figure 6: Using the natural language query feature.

This feature uses advanced language processing techniques to understand and process your culinary queries effectively.

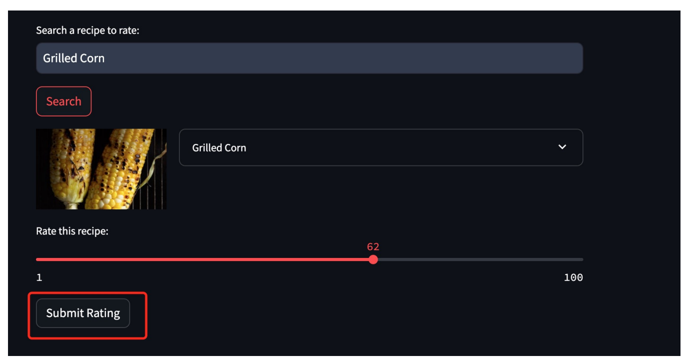

Figure 7: Rating interface for user feedback on recipes.

7.4 Rating and Feedback System

Your feedback is invaluable in improving recipe recommendations. Rate the recipes you try to help refine and personalize future suggestions:

(1) After preparing a dish, go to the recipe's page on our system.

(2) Provide a rating based on your satisfaction with the dish.
(3) Optionally, add comments to help others understand your experience and to provide feedback for system improvements.

Ratings influence how recipes are presented and recommended to you and other users, ensuring that the most liked dishes are easier to find.

Summary: By engaging with these features, you can enhance your cooking experience, discover new favorite recipes, and make meal preparation easier and more enjoyable. Follow these guidelines to maximize the benefits of our Recipe Recommendation System.