

PantryPal

Abstract

This project aims to develop an AI-powered recipe generator that suggests recipes based on available ingredients, kitchen tools, and time constraints. Data will be sourced from online recipe databases and structured ingredient lists. The system will use Natural Language Processing (NLP) techniques with rule-based filtering to generate recipes that are practical, coherent, and user-friendly. Evaluation metrics we will aim to use include BLEU scores for textual coherence, user ratings for satisfaction, and perplexity to assess the models's fluency. A hybrid approach will be adopted to ensure high-quality generation, leveraging both data-driven and rule-based methodologies.

Background

Automated recipe generation has been studied in the fields of Natural Language Processing (NLP) and recommendation systems. Prior work includes structured dataset approaches, such as RecipeNLG, which provides a large amount of recipes for semi-structured text generation [1]. Deep learning-based text generation has shown promise in generating coherent recipe output and contextually appropriate recipes [2]. For example, Kiddon et al. (2016) demonstrated the effectiveness of neural checklist models in generating globally coherent text, which is highly relevant for recipe generation tasks.

In addition to data-driven approaches, rule-based systems have been widely used to enforce practical constraints in recipe generation. For instance, Teng et al. (2012) explored rule-based methods for recipe recommendation, focusing on ingredient compatibility and dietary restrictions [3]. These systems are particularly useful for filtering unrealistic combinations of ingredients and ensuring that generated recipes follow user constraints, such as dietary preferences or time limitations.

Hybrid approaches, which combine data-driven models with rule-based systems, have also been applied specifically to recipe generation. For example, Chow et al. (2023) proposed a hybrid system for recipe recommendation that integrates collaborative filtering (a data-driven model) with rule-based filtering to improve recommendation quality and adaptability to user preferences [4]. Hybrid systems highlight the increasing need to integrate dietary preferences, nutritional needs, and ingredient compatibility into food recommender systems, making them more personalized and practical for users.

The motivation for this project comes from the growing interest in personalized meal planning. By utilizing machine learning, this project aims to create a system that not only generates recipes but also adapts to user constraints, like dietary restrictions and time limitations. The project

builds on prior research in recommendation systems, while addressing the practical concerns of real-world recipe generation.

Problem Statement

The challenge is to generate contextually relevant and coherent recipes given user constraints. The problem is quantifiable through recipe coherence metrics like BLEU scores, and user validation can be measured through ratings. The project's performance can be assessed through perplexity, to evaluate the model's confidence in generating coherent text, and user feedback. The problem can be reproduced using publicly available recipe datasets and standardized evaluation metrics.

Data

1. Recipe Ingredients Dataset [5]

- **Size:** Large-scale dataset with diverse recipes
- **Key Features:** Ingredient-centric data with metadata
- **Possible Usage:** ingredient standardization and compatibility analysis

2. Food.com Recipes and Interactions [6]

- **Size:** 180K+ recipes with 700K+ user interactions
- **Key Features:** User reviews, ratings, and feedback
- **Possible Usage:** Model evaluation and personalization analysis

3. Auntie's Recipes [7]

- **Size:** Community-drove collection of recipes
- **Key Features:** Unique recipes
- **Possible Usage:** Supplementary dataset to enhance the diversity

4. Recipe Box [8]

- **Size:** Curated set of recipes with structures metadata
- **Key Features:** Open-Source, user-contributed recipes with a variety of cuisines
- **Possible Usage:** Additional data for increasing recipe variety

Proposed Solution

A transformer-based text generation model (e.g., GPT-3 fine-tuned on recipe data) will be used to generate structured and coherent recipes. NLP preprocessing will leverage [spaCy](#) and [NLTK](#) to clean and standardize ingredient names and instructions. Rule-based logic will filter out unrealistic ingredient combinations and ensure adherence to dietary restrictions. An attention-based approach may be incorporated to enhance the structuring of generated recipes.

The final pipeline will take user inputs on available ingredients, kitchen tools, and time constraints to produce feasible and contextually relevant recipes.

Evaluation Metrics

- **BLEU Score:** Measures text coherence compared to the reference recipes
- **Perplexity:** Assesses language model fluency
- **User Ratings:** Gathers real-world feedback on recipe usability

To quantify the performance of both the benchmark model and the proposed solution, we will use a combination of BLEU score, perplexity, and user ratings. The BLEU score measures the similarity between the generated recipes and human-written recipes by comparing sequences of words. For instance, a BLEU score close to 1 indicates high similarity to human-generated recipes, providing a measure of recipe coherence that is quantitative. Perplexity determines the model's confidence in generating coherent text, with lower values closer to 0 indicating better performance. Furthermore, user ratings will be collected to assess the practicality of the generated recipes giving us real-world feedback on the system. These metrics are mathematically quantifiable and work together to provide a comprehensive evaluation on the system's performance.

Ethics & Privacy

The project has some ethical concerns that must be addressed. The system must account for dietary restrictions and allergens to avoid harmful recommendations, which could cause serious health implications for users. The model will include filtering options for dietary preferences to combat this. Additionally, the collection and use of user-provided data, could potentially reveal sensitive information about individuals, including health conditions. To mitigate privacy risks, all user data will be anonymized and stored securely, with consent for data usage.

References:

- [1] Bień et al., INLG 2020. RecipeNLG: A Cooking Recipes Dataset for Semi-Structured Text Generation
- [2] Kiddon, C., et al. (2016). Globally Coherent Text Generation with Neural Checklist Models. NAACL.
- [3] Teng, C.-Y., et al. (2012). Recipe Recommendation Using Ingredient Compatibility.

[4] Chow, Y.-Y., et al. (2023). Food Recommender System: A Review on Techniques, Datasets, and Evaluation Metrics.

[5] [Kaggle: Recipe Ingredients Dataset](#)

[6] [Kaggle: Food.com Recipes and User Interactions](#)

[7] [GitHub: Auntie's Recipes](#)

[8] [Eight Portions Recipe Dataset](#) ([GitHub Repository](#))

Dataset possibilities:

1. <https://www.themealdb.com/>
2. <https://www.kaggle.com/datasets/kaggle/recipe-ingredients-dataset>
3. <https://www.kaggle.com/datasets/shuyangli94/food-com-recipes-and-user-interactions>
4. <https://github.com/user24/auntiesrecipes>
5. <https://eightportions.com/datasets/Recipes/#fn:1>
 - a. <https://github.com/rtlee9/recipe-box>
6. <https://opendata.stackexchange.com/questions/4283/open-downloadable-recipe-database>
 - a. This is like a list of hell
- 7.

Example githubs:

<https://github.com/kolibri753/recipe-generator-powered-by-ai>

<https://github.com/Jeanette-Ralph/Recipe-Generator> (in java)

https://github.com/dhirajpatra/recipe_generator (more complicated - input time, tools, chef level)

<https://github.com/elenalape/chef-ai>

Proposed Solution (this might not work?)

1. Data Preprocessing

- Standardize ingredient names and measurement units
- Tokenize recipe steps using NLP tools (spaCy, NLTK)
- Remove duplicate recipes and incorrect entries

2. **Model Training**

- Fine-tune a transformer-based language model (e.g., GPT-3, T5) on TheMealDB, Kaggle datasets, Auntie's Recipes, and Recipe Box
- Incorporate rule-based logic to filter out unrealistic ingredient combinations
- Implement attention mechanisms to ensure logical sequence in recipe steps

3. **Recipe Generation Pipeline**

- User inputs available ingredients, kitchen tools, and time constraints
- Rule-based system filters feasible ingredient combinations
- Transformer model generates recipe text with structured output
- Post-processing ensures consistency in instructions and ingredient use

4. **Evaluation Metrics**

- **BLEU Score:** Measures text coherence compared to reference recipes
- **Perplexity:** Assesses language model fluency
- **User Ratings:** Gathers real-world feedback on recipe usability