AI-Powered Food Discovery and Recommender App

Authors: Abhishek Tibrewal (MT23009), Aruba Sood (MT23022), Mayank (MT23045), Purbasha Barik (MT23068), Mehakdeep Kaur (MT23126), Rajith Ramchandran (MT23132)

Problem Statement

Develop an AI-powered Food Discovery and Recommender Application that revolutionizes the culinary experience by leveraging image recognition technology. The application should be capable of scanning an image of a food item, accurately identifying it, and providing a comprehensive suite of features including:

- **Complementary Food Recommendations:** Suggesting additional food items that pair well with the identified dish to enhance the dining experience.
- **Ingredient List Generation:** Displaying a list of ingredients required to prepare the identified dish.
- **Recipe Provision:** Offering a detailed recipe for creating the dish at home.
- **Instructional Video Access:** Providing a direct link to a YouTube video tutorial for step-by-step cooking instructions.

The application aims to cater to food enthusiasts who wish to explore new flavors, aspiring chefs seeking culinary inspiration, and anyone interested in learning about diverse cuisines through an interactive and informative platform. The challenge lies in creating an intuitive and user-friendly interface that delivers accurate food identification and relevant recommendations, thereby enriching the user's gastronomic journey.

Progress

We have made some substantial advancements in our Project. We have done thorough literature review, dataset construction, and preprocessing efforts. The literature review phase provided a comprehensive understanding of existing approaches in culinary applications, identifying gaps and opportunities for innovation. Drawing upon insights from diverse sources, we formulated a clear roadmap for the development of our artificial intelligence-powered food discovery and recommender app.
Subsequently, we have constructed a significant dataset representative of real-world culinary scenarios. This dataset encompasses a varied range of food items captured from various perspectives and conditions to facilitate robust training of image recognition algorithms. This dataset consists of a total of 3,160 images from 63 different classes of food, with each class containing 50 images. Moreover, several preprocessing steps were implemented to enhance the dataset's quality and ensure compatibility with our proposed methodologies. These steps include resizing, grayscaling, flipping, brightness & contrast adjustments, rotation etc.

Overall, the progress achieved in these initial phases sets a solid foundation for the subsequent stages of app development and evaluation, positioning us favorably to deliver a cutting-edge solution to address the identified challenges in culinary applications.

Detailed Literature review:

1) *<u>Cooking Recipe Analysis based on Sequences of Distributed Representation on Procedure Texts and Associated Images</u>* *<u>[Akari Ninomiya; Tomonobu Ozaki; Nihon University Tokyo, Japan]</u>*

The paper introduces a method to represent cooking recipes using distributed representations from text and associated images. By analyzing the dissimilarity among these sequences, the study aims to provide a precise representation for cluster analysis. The approach is evaluated on four dishes, demonstrating its effectiveness in understanding recipe relationships based on DTW distance.

Dataset used:

- The paper utilized the "Cookpad dataset" provided by Cookpad Inc. via the IDR Dataset Service of the National Institute of Informatics.
- The Cookpad dataset likely contains a collection of cooking recipes for analysis in the research.
- Recipe images from the Cookpad Image Dataset were also used in the study.
- The Cookpad Image Dataset is mentioned to be an image collection serving as infrastructure for food research.

These datasets were essential for conducting the analysis and experimentation in the research on cooking recipe analysis based on distributed representations of text and images.

Techniques Employed:

The techniques used in the paper for analyzing cooking recipes can be summarized as follows:

- Representation Learning:
    - Utilizing BERT for textual representation: Cooking steps written in recipe text are processed using BERT to create a distributed representation sequence.
    - Leveraging VGG16 for image representation: Sequences of associated images during cooking are analyzed using the VGG16 convolutional neural network to generate another distributed representation sequence.

- Cluster Analysis:
    - Dynamic Time Warping (DTW) distance: Relationships among recipes are examined by performing cluster analysis based on the standard DTW distance metric applied to the distributed representation sequences of cooking steps and associated images.

These techniques aim to provide a detailed and precise measurement of relevance among recipes by representing them using distributed representations derived from both textual and visual data and then analyzing these representations through cluster analysis based on DTW distance.

2) *Food Recipe Alternation and Generation with Natural Language Processing Technique [Yuran Pan,Qiangwen Xu; Yanjun Li: Dept. of Computer and Information Sciences Fordham University New York, U.S.A.]*

The paper "Food Recipe Alternation and Generation with Natural Language Processing Techniques" explores the application of natural language processing (NLP) techniques in the culinary domain. It focuses on assisting users in selecting alternative ingredients and recipes by utilizing word embedding and similarity measurements. The research project employs N-gram and neural network models to generate new recipes with authentic flavors of various cuisine styles. The ultimate aim is to help individuals facing ingredient limitations and encourage culinary exploration. The study also discusses dataset collection, experimental results, and future work, such as expanding the dataset and collaborating with chefs for enhanced recipe interpretation.

Dataset Used:
The dataset used in the paper consists of 3433 recipes across 15 different cuisine styles, collected from the Spoonacular website.

- Each recipe includes various features such as cuisine type, title, ingredients, process, diets, and nutrition information. After cleaning and preprocessing the data, a new feature named "process" was generated based on the cooking instructions. Additionally, an ingredient-process pairing was added to the dataset as an extra feature.
- The dataset analysis includes studying ingredient distribution among cuisine styles, with American style having the highest number of ingredients (683) and African style having the lowest (364). Word clouds were created to visualize ingredient frequency, showing common ingredients used across all cuisine styles.

The dataset serves as the foundation for applying natural language processing techniques to suggest ingredient alternatives and generate new recipes with authentic flavors.

Techniques Employed:
Word Embedding: Using word embedding to represent ingredients and measure similarity between them.

- N-gram Language Model: Employing N-gram language model to generate new recipes based on training data.
- Neural Network Models: Specifically, Recurrent Neural Network (RNN) with Long Short-Term Memory (LSTM) layer for language modeling.
- Similarity Measurement: Calculating cosine similarity between ingredient embedding vectors to determine ingredient and recipe similarities.
- Data Preprocessing: Traditional NLP techniques like tokenization, stemming, and stop words removal for dataset preparation. These techniques are applied to assist users

in selecting alternative ingredients, generating new recipes, and exploring culinary options.

3) *Food Image Classification with Convolutional Neural Networks [Malina Jiang Department of Computer Science Stanford University]*

This paper explores using CNNs to classify food images, aiming to enhance food experiences and aid in dietary choices. It compares training CNNs from scratch to transfer learning with pre-trained weights, achieving 61.4% accuracy and 85.2% top-5 accuracy. The best model is a pre-trained InceptionV3 with gradually unfrozen layers during transfer learning. Future work includes optimizing hyperparameters and adding features like bounding boxes to boost classification accuracy.

Dataset used:

- Dataset: Food-101 dataset with 101,000 images across 101 food classes.
- Test Images: 250 manually reviewed test images per class, with 750 intentionally noisy training images.
- Challenges: Diverse image characteristics (lighting, coloring, size), potential mislabeled images, and similar food categories.
- Preprocessing: Images normalized and resized to 128x128 or 256x256, with data augmentation techniques like rotation and flipping.
- Transfer Learning: Custom model preprocessing functions based on original model papers used during transfer learning.
- Significance: Food-101 dataset offered a diverse training and testing environment for convolutional neural network models in food image classification .

Techniques Employed:

- Data Augmentation: Introducing transformation functions to combat overfitting and enhance training data diversity.
- Transfer Learning: Leveraging pre-trained models (VGG16, ResNet50, InceptionV3) from ImageNet, replacing top layers for Food-101 classification.
- Model Architecture Optimization: Experimenting with various architectures, including baseline and Lu's proposed model, to enhance classification performance.
- Image Preprocessing: Normalizing and resizing images for model input, utilizing custom preprocessing functions during transfer learning.
- Loss Function: Employing categorical cross-entropy as the loss function for all models to measure class label prediction variance.

4) *A Cooking Recipe Recommendation System with Visual Recognition of Food Ingredients [Keiji Yanai, Takuma Maruyama and Yoshiyuki Kawano The University of Electro-Communications, Tokyo, Japan]*

The research paper presents a Cooking Recipe Recommendation System for smartphones, allowing users to access recipes by pointing their cameras at ingredients. It addresses the challenge of recipe access while shopping and utilizes object recognition technology for real-time suggestions. The system employs a color-histogram-based approach for image representation, simplifying the cooking decision-making process and offering a user-friendly solution for quick recipe recommendations based on recognized ingredients.

Dataset Used:

The data set used in the experiments for this paper consisted of 30 kinds of food ingredients. These ingredients were collected through short videos recorded at grocery stores in Tokyo. Each ingredient category had 10 short videos recorded at 25 frames per second with a VGA resolution (640x480). The data set included various types of ingredients such as fish, meat, vegetables, and fruits. Here is a list of the 30 kinds of food ingredients in the data set:

- Fish (5): tuna, squid, octopus, shrimp, salmon
- Meat (5): beef, pork, chicken, minced meat, sausage, ham
- Vegetable (13): mushroom, potato, eggplant, carrot, radish, tomato, cucumber, cabbage, green onion, onion, Chinese cabbage, lettuce, Shiitake mushroom
- Fruit (6): apple, strawberry, pineapple, orange, banana, grapefruit
- This diverse data set was used for object classification performance evaluation in the experiments conducted for the system

Techniques Employed:

This paper utilized several key technologies to enable real-time object recognition of food ingredients and recommend cooking recipes based on the recognized ingredients. Here are the technologies employed in the system:

- Image Features:
  - SURF (Speeded-Up Robust Features): Used for invariant local feature extraction for scale, rotation, and illumination changes.
  - Color Histograms: Grid-based color histograms extracted from images in RGB, HSV, and Lab color spaces to capture important color information.

- Image Representation:
  - Bag-of-Features: Conversion of local color features and SURF features into bag-of-feature representation for classification.

- Image Classifier:
  - Linear Kernel Support Vector Machine (SVM): Employed for real-time recognition with low computational cost. The system used a linear SVM classifier with the one-vs-rest strategy for recognizing 30 kinds of food ingredients.

- Mobile Device Platform:
  - Android-based Smartphones: The system was designed to run on consumer smartphones, enabling users to perform real-time object recognition by pointing the device's camera at food ingredients.

- Data Collection:
    - Short Videos: Data set collection involved recording short videos of 30 food ingredient categories at grocery stores in Tokyo, with each video lasting about 5 seconds at 25 frames per second.

These technologies were integrated to create a user-friendly and efficient system that could recommend cooking recipes based on visual recognition of food ingredients in real-time

5) *Recipe2Vec: Multi-modal Recipe Representation Learning with Graph Neural Networks [Yijun Tian, Chuxu Zhang, Zhichun Guo, Yihong Ma, Ronald Metoyer, Nitesh V. Chawla]*

Recipe2Vec is a novel model for multi-modal recipe representation learning that integrates visual, textual, and relational information using Graph Neural Networks. It outperforms existing methods by effectively capturing the nuances of recipe data. The Large-RG recipe graph dataset facilitates graph-based food studies and enhances the model's performance.

Dataset Used:

- Dataset: Utilized Recipe1M+ dataset for learning cross-modal embeddings for cooking recipes and food images.
- Content: Contains relational information and multi-modal data (images, text) essential for the study's multi-modal recipe representation learning.
- Purpose: Designed to facilitate learning cross-modal embeddings for recipes and images, enhancing the understanding of food content and relationships.
- Significance: Enables comprehensive exploration of multi-modal data to improve recipe representation and recommendation models.

Techniques Employed:

- Graph Neural Networks (GNNs): Proposal of Recipe2Vec, a novel model utilizing GNNs for multi-modal recipe embedding.
- Multi-modal Information Integration: Focus on combining visual, textual, and relational data for enhanced recipe embeddings.
- Adversarial Attack Strategy: Introduction of a strategy to ensure stable learning and boost model performance.
- Node Classification: Design of a joint objective function for node classification and adversarial learning to optimize the model.
- Multi-view Neighbor Sampler: Introduction of a sampler to capture local and high-order information from different types of neighbors.
- Feature-based Adversarial Attack: Utilization of a feature-based strategy to improve stability and performance of the model.

Gaps/Limitations of the Literature Review

Based on our literature review we have managed to find some potential gaps or areas where the proposed solution could introduce novelty:

- Integration of Image Recognition and Natural Language Processing: While the papers discuss the use of image classification and natural language processing separately , , there is an opportunity to combine these modalities in a unified system. By integrating image recognition algorithms with natural language processing techniques, the app could provide a more comprehensive analysis of food items by considering both visual and textual information simultaneously.

- User Interaction and Personalization: The papers mention the use of language models for recipe generation and analysis ,but there is room to enhance user interaction and personalization. By incorporating a chatbot powered by natural language processing, the app could offer personalized recommendations, answer user queries, and engage users in interactive conversations to better understand their preferences and dietary restrictions.

- Evaluation Metrics and User Feedback: While the papers discuss experimental results and model performance evaluation , there is an opportunity to include a more robust evaluation framework. By incorporating user satisfaction surveys, feedback mechanisms, and usability testing, the app can gather valuable insights on user experience, satisfaction levels, and areas for improvement.

- Recommendation System Enhancement: The papers touch upon recommendation systems for recipes based on user preferences , but there is potential to enhance the recommendation algorithms.

In summary, while the Reviewed Papers provide valuable insights into recipe analysis, generation, and representation learning, there are opportunities to enhance the proposed solution by integrating image recognition with natural language processing, focusing on user interaction and personalization, improving evaluation metrics and user feedback mechanisms, and enhancing the recommendation system algorithms.