Recipe Recommender System with Ingredient Object Detection

Shuangyue Cheng, Shuyan Lin, Shiyi Wang, Haoran Zhao, Taichang Zhou
School of Computer Science, Georgia Institute of Technology, Atlanta, Georgia, USA
{katcheng, slin915, swang793, hzhao353, tzhou915}@gatech.edu

KEYWORDS

Content-based Recommender System, Web Crawler, Domain-Specific Object Detection, Deep Learning based Ranking

1 Motivation

Food, as such an essential component of human experience, significantly influences our feelings and health. With the development of mass transportation and the global internet, people from different parts of the world are becoming more connected. Food has demonstrated a remarkable potential to bring cultures together. According to the National Restaurant Association, about 80% of surveyed participants eat at least one "ethnic" cuisine each month. As an evolving trend in American diets, two-thirds eat a wider variety of cuisines than they did 5 years ago (NRA, 2015).

However, multi-language recipe search and recommendation remained an under-explored field (Towmey et al., 2020). Imagine you have the napa cabbage at hand, and you would like to have some Korean recipe on it, so you google "napa cabbage Korean recipe." None of them looks like the one you have tried in the Korean restaurant. Then you think it may be better to search for it with its Korean name. The google translation gets you the answer "배추" (baechu). It doesn't look helpful because it commonly directs you to a more Korean translation needed page. In the case of cooking authentic ethnic food from a non-English region, you can pretty frequently find yourself stuck because you cannot utilize a foreign language recipe website efficiently. The easy answer for this story is "kimchi." Still, we can't remember every ethnic food dish name, not to mention those authentic ones with no English name. In other cases, users may want to try some genuine cuisine. Searching in English usually results in a higher probability of getting American-styled food recipes.

This project aims to create a content-based filtering recipe recommender system that provides global recipes based on the user ingredient input. There's no need for the user to know an ingredient's name in another language with an ingredient object detection module. A user only needs to take a picture of the ingredients and submit it to our website. The system will recognize the item and output a list of most-matched global recipes in English. Text input for ingredients in English is also supported if the user prefers to type. With the steps, images, and possibly a video link, users can efficiently obtain authentic dish recipes with no language barriers.

2 Related Work

2.1 Static and Dynamic Page Web Crawler

Web crawlers can be categorized into three types: general purpose crawling, distributed crawling, and focused crawling. While general purpose crawling and Distributed crawling mainly focus on crawling quantities and performance, focused crawling only collects pages on a predefined topic (Abu Kausar et al., 2014), which in this project is several selected recipe websites. Target webpages can be categorized into two structure types: static pages and dynamic pages. While static page crawling can simply capture the paragraphs and titles with the Html elements, dynamic page crawling must execute JavaScript to obtain the content. The JavaScript evaluation significantly slows down the page loads, (Wang., 2016) which in turn slows down our information retrieval in crawling. Depending on the target website structure, a focused web crawler will be launched to crawl a combination of static and dynamic pages, while static ones are more preferred.

2.2 Content Based Recommender System

There are two types of traditional recommender system algorithms: content-based and collaborative filtering. Content-based filtering uses item features to recommend items like what the user likes, based on their previous actions or explicit feedback. On the other hand, collaborative filtering simultaneously uses similarities between users and items to provide recommendations (Google Developers, 2021). Due to the constraint of our dataset where we do not have access to the user ratings, we choose to use a content-based filtering technique. Existing literature has demonstrated implementation details of content-based filtering, such as stemming and lemmatization (Stanford NLP Group, 2009), nltk, the Term Frequency-Inverse Document Frequency (TF-IDF), and cosine similarity (Swathi et al., 2017).

2.3 Image Recognition

Most of the datasets contain food specific to certain regions, such as ChineseFoodNet. It's also important to note that region-specific food datasets can undermine the accuracy. Therefore, we need more generic food datasets. When picking the dataset, it should be kept in mind. Feature extraction is another aspect of food recognition. It should identify characteristics of food such as color, shape, and texture. There are two feature extraction techniques: handcrafted and deep visual features. By combining the two, it also results in greater performance. Handcrafted aims to identify relevant features of objects. Deep model is automatic feature extraction. Regarding the deep visual, there are CNN, DCNN, and Inception-v3. DCNNs are suitable for large-scale object recognition. It's more breath, and Inception-v3 is more depth. It has higher accuracy but is computationally expensive. CNN has more outstanding performance for extracting feature descriptors. Using DCNN, the accuracy of the food dataset is around 70% (Tahir, Loo, 2021).

3 System Design

The goal of our recipe recommender system is to output a list of recipes given user ingredient inputs in both image and text. We will implement a web crawler, a content-based filtering recommender system, and an object recognition system to achieve the goal.

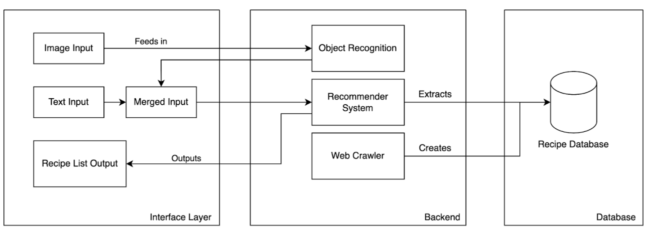

Figure 1: Overall System Architecture Design

3.1 Crawler Design

We would like to first focus our recipe database on three areas: American, Chinese, and Korean food. We found an excellent open resource dataset for American food of 180K+ recipes from Food.com on Kaggle. We also selected two popular Chinese and Korean recipe websites as the seed URLs. Once it gets the URL, it downloads the web data into the local file system and extracts related valuable links in the page. With a combination of first depth search and breadth depth search algorithm, it should crawl as many non-duplicate pages as possible in the specified domains. Before storing information in our database, we launch a translation service to convert keywords into English. We will establish a consistent database with global recipes after cleaning and verification.

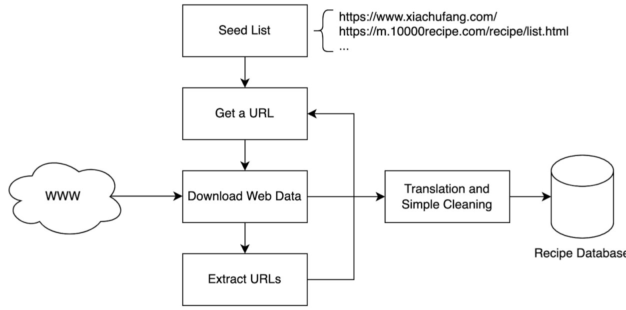

Figure 2: Crawler Architecture Design

Table 1. Recipe Data Source Overview

Recipe	Data Source
American	https://www.kaggle.com/shuyangli94/food-com-recipes-and-user-interactions
Chinese	https://www.xiachufang.com/ https://home.meishichina.com/recipe.html
Korean	https://m.10000recipe.com/recipe/list.html http://ottogi.okitchen.co.kr/m_bestList.asp

3.2 Recommender System Design

The overall design of our content-based filtering recommender system consists of four stages: data preprocessing, feature extraction, similarity calculation, and ranking. The first stage is data preprocessing. In this stage, we will remove irrelevant information from the scraped ingredient list with nltk. Besides, we need to deal with different forms of the same vocabulary, such as singular vs. plural. With stemming and lemmatization, we will address the cases above. The second stage is feature extraction. We need to embed our ingredients into vectors to make computations. In this case, we can use TF-IDF. The third stage is similarity calculation. In this case, we can apply cosine similarity. Finally, we can supply a ranked output list.

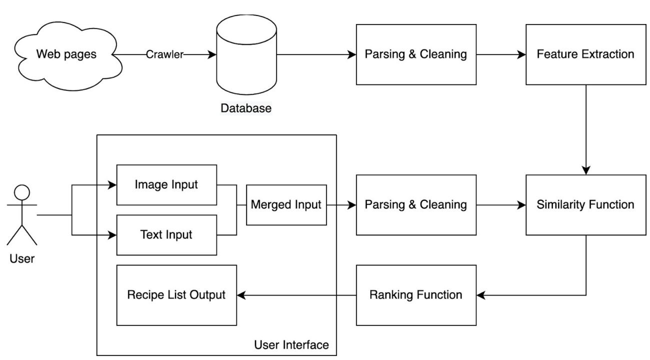

Figure 3: Recommender System Architecture Design

3.3 Image Recognition Design

We aim to implement an image recognition system to allow users to input and output a list of ingredients from the picture. A convolutional neural network (CNN) can be advantageous because as the dataset grows, the accuracy increases. There are three types of layers: convolutional, pooling, and fully connected. CNN takes the raw pixel of the image and extracts features through layers by reducing the dimension of features and inferring what the food item is. This is done by extracting features using filters. CNN can learn the optimal values for the filter to extract the most meaningful. After convolution, ReLU transformation is applied and then pooled to reduce the dimension of features. Finally, there is a fully connected layer for classification. Fully connected means every node in the first layer is connected to a node in the second layer. It outputs the probability for each classification label.

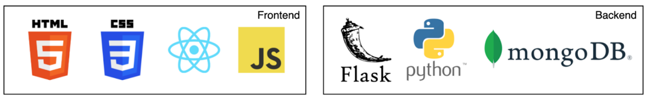

Figure 4: Framework, Language, and Technology Choices

4 Division of Work

We assigned our specific responsibilities to each team member utilizing a Gantt Chart. The schedule details are displayed below.

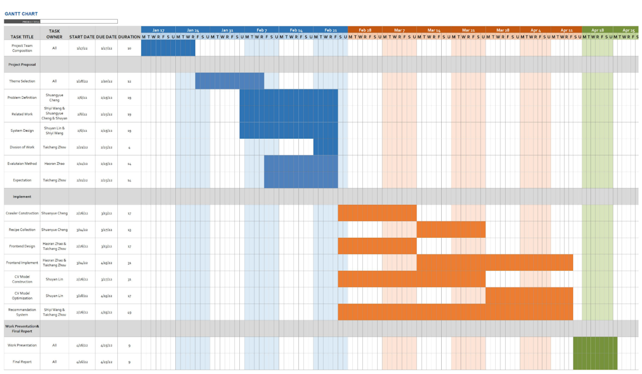

Figure 5: Gantt Chart for Work Division

5 Deliverables and Evaluation

We aim to implement a web-based application where a user can upload an ingredient picture and get the corresponding recipe list and an URL of video tutorials if available. We expect our application to work for all kinds of cuisine but may change the scope based on our progress.

The accuracy of the image recognition system and recipe crawler will be two key factors to be evaluated. We will examine the result of ingredient object recognition by uploading a real-life picture outside the dataset. In terms of analytics, we will apply k-fold cross-validation to investigate the accuracy of our system. To test the recipe crawling result, we will compare the result from our crawler with our manual crawling. We will examine if our web crawler can achieve satisfactory accuracy.

6 Future Work

6.1 Recommender System Extensions

Deep Learning-Based Ranking: the ranking system will use the NLU model with stemming and inverse document frequency to extract ingredients from the recipe. Before ranking, we will build a normalized vector of ingredients of each recipe. The algorithm to find the most similar recipes is not determined now. We will try algorithms like KNN and cosine similarity.

Enhanced Data Cleaning: for those recipes with little difference, we will detect them with cosine similarity of ingredient vectors scaled by used amount. The algorithm will list all similar ingredients and check manually to remove real duplicates.

6.2 Application Feature Extensions

Crawled Sourcing Recipe Categories: the crawled recipe will be divided into complete and overlapping categories so that the user can only look for specific recipes.

User Profile Personalized Recommendation: the search history, ratings, and comments will be stored to build a more focused on recent history to develop a personalized recommendation system. This also gives a weight list that will be considered during ranking.

REFERENCES

[1] Google Developers. 2021. Content-based Filtering | Recommendation Systems. *Google Developers*. Retrieved February 23, 2022 from https://developers.google.com/machine-learning/recommendation/content-based/basics
[2] Abu Kausar, V. S. Dhaka, and Sanjeev Kumar Singh. 2013. Web Crawler: A Review. *International Journal of Computer Applications* 63, 2 (February 2013), 31–36. DOI:https://doi.org/10.5120/10440-5125
[3] National Restaurant Association. 2015. Global Palates: Ethnic Cuisines and Flavors in America (PDF). *National Restaurant Association*. Retrieved February 24, 2022 from https://shop.restaurant.org/Global-Palates-Ethnic-Cuisines-and-Flavors-in-America-PDF
[4] Stanford NLP Group. 2009. Stemming and lemmatization. *Stanford.edu*. Retrieved February 23, 2022 from https://nlp.stanford.edu/IR-book/html/htmledition/stemming-and-lemmatization-1.html
[5] Swathi, S. Devi, Dr Shelbi Joseph, and Seetha Parameswaran. 2017. Various Methods of Using Content-Based Filtering Algorithm for Recommender Systems. *www.semanticscholar.org*. Retrieved February 23, 2022 from https://www.semanticscholar.org/paper/Various-Methods-of-Using-Content-Based-Filtering-Swathi-Devi/c972cd00cda0ae52d0347f0f912170fdeb6a3b65
[6] Ghalib Ahmed Tahir and Chu Kiong Loo. 2021. A Comprehensive Survey of Image-Based Food Recognition and Volume Estimation Methods for Dietary Assessment. *Healthcare* 9, 12 (December 2021), 1676. DOI:https://doi.org/10.3390/healthcare9121676
[7] Niall Twomey, Mikhail Fain, Andrey Ponikar, and Nadine Sarraf. 2020. Towards Multi-Language Recipe Personalisation and Recommendation. *Fourteenth ACM Conference on Recommender Systems* (September 2020). DOI:https://doi.org/10.1145/3383313.3418478
[8] Xiao Wang, Arvind Krishnamurthy, and David Wetherall. 2016. *Speeding up Web Page Loads with Shandian Speeding up Web Page Loads with Shandian*. Retrieved February 24, 2022 from https://www.usenix.org/system/files/conference/nsdi16/nsdi16-paper-wang-xiao-sophia.pdf
[9] Zhongwei Xie, Ling Liu, Lin Li, and Luo Zhong. 2021. Learning Joint Embedding with Modality Alignments for Cross-Modal Retrieval of Recipes and Food Images. *Proceedings of the 30th ACM International Conference on Information & Knowledge Management* (October 2021). DOI:https://doi.org/10.1145/3459637.3482270