

Project 4: Natural Language Processing and Unsupervised Learning
Recipe Recommender
Flora Xinru Cheng
November 3, 2019

Domain

I learned to cook during college via online recipes in both English and Chinese. At first I tried to recreate recipes from my childhood with alternative ingredients, then I started to expand my repertoire with recipes from different cuisine types. Online recipes are very convenient to follow, but the hard part is not knowing what to search for.

As online grocery/meal delivery services such as Blue Apron and Hellofresh become more popular, there is a need for variety in recipes based on a few given ingredients (Hellofresh: "...the world's leading meal kit provider, we delivered 65.6 million meals to 2.48 million active customers worldwide in Q1 2019, in 11 countries across 3 continents.").

As someone who loves to travel as well as exploring foods from different cultures, I often notice similarities from recipes belonging to different cuisine types. I would like to make a tool to help fellow foodies and travellers like me discover recipes they might enjoy from other cultures more easily. They could use it to order their meals when trying out a foreign restaurant, with more confidence that they will like the dish being served. This tool can also help add more variety and excitement for cooking with a few basic ingredients at home.

Design

The end goal of this project would be a recipe recommender in the form of a working Flask app, that is ideally also mobile-friendly. I want to turn data into something my family and friends could use to find recipes based on a few ingredients. Having a web app as the end goal also helps with measuring progress, since it would be difficult to measure the quality of recommendations, or know when to stop tuning the system.

I am aware that making my first web app and deploying it might take up a significant amount of time, and I would likely have to sacrifice some functionality of the recommendation system as well as the app to get to a working product (I am also open to the possibility of expanding this into my passion project).

Data

Because recipes are time-insensitive, I will be using downloaded data scraped and uploaded by others. The best data source I found so far is recipe box (<https://eightportions.com/datasets/Recipes/#fn:1>), which has ~125,000 scraped recipes from food websites (Foodnetwork.com, Epicurious.com, and Allrecipes.com). The dataset is 50.9MB, in JSON format. Each recipe entry contains **a recipe title, a list of ingredients and measurements, instructions for preparation**, and many also have a picture of the resulting dish.

Although the Kaggle What's Cooking (<https://www.kaggle.com/c/whats-cooking/data>) dataset is the smallest and cleanest food dataset I've found (and initially seems great for an mvp), it does not have names for recipes. The only features are **recipe ids, cuisine types, and ingredients** (The source for that data (Yummly.com) has an API but it shut down in September). But because it comes with pre-labelled cuisine types which recipe box lacks, it might still be useful for cuisine classification.

Minimum Viable Product

An mvp for this project would involve selecting and cleaning data (likely from just one of the scraped websites), classifying recipes according to cuisine type, using topic modelling or some other method to extract the main ingredients from each recipe, then building a simple recommendation system based on recipe similarity (shared ingredient topics).

To expand on this I would include more data from other websites, and improve the recommendation system. I am also interested in clustering the recipes by cuisine type, as well as making recommendations based on similarities in flavors (and even expanding to recipe names from other languages), but probably wouldn't have time due to needing to make the web app.

Tools

Python, pandas, NumPy, scipy, matplotlib, seaborn, Tableau
scikit-learn, pickle
nltk, Gensim, spaCy
Flask, HTML, CSS, possibly also some basic JavaScript
GitHub, cookiecutter (datasciencemvp template)
AWS, Jupyter Notebook, Visual Studio Code