Recipe Recommender

UBC-CPEN291/project-team-apatosaurus

Harrison Mitgang (hmitgang), Benjamin Ren (boweiren), Alina Shabanova (alinas2000), Jane Wu (jwu611)

Current state

We have collected all the appropriate data that will be used to test and train the ML components. This includes the food image dataset, as well as the recipe, ingredients, and ratings dataset. The next steps are to train the classification model as well as the recommender system. We still need to implement all parts of the website.

Feature Changes

As this is our first milestone, we slightly amended the proposal after group discussion.

For Data/Sources, we will:
- Use Bing API *or* Bing downloader script (python package).
- Only provide image support for a subset of ingredients:
  - Meat (beef, salmon, chicken)
  - Vegetables (broccoli, cabbage, carrot, celery, corn, cucumber, eggplant, green bean, bell pepper, olive, onion, potato, spinach, tomato, lettuce)
  - Fruit (apple, avocado, banana, lemon)
  - Others (bread, cheese, mushroom, egg, pasta, rice)
- Gather ~300 images per ingredient in the subset.
- Remove recipes that don't contain *any* of the supported ingredients.

For the Non-ML Component:
- Instead of using PythonAnywhere (which has a few limitations on outbound Internet requests), we will use the free Heroku package.

For Milestones:
- For milestone 2, we will work on the classification model as well as begin on the recommender model, with two people on each model.
- For milestone 3, we will perfect the two models and start working on the website.

Challenges

- We are expecting to use dataset images of the raw ingredient, so some images that were obtained with an original search did not meet the requirements of being appropriate to be used as part of the dataset, such as images of the ingredient in cooked dishes. This required some team members to filter through them and adjust the search name to obtain the most relevant images.

- Some ingredients cannot be searched easily with a single label, so it may require some searching technique to finally get the images meeting the requirement.
- There are *many* ingredients in the Food.com dataset, so filtering through them to match similar ingredients is difficult. Since generating our dataset from Food.com can be systematically replicated, if we find that some labels that we have not already thought of should be combined, it is easy to do so in the `data_gathering.ipynb` notebook.

Member Tasks

- Alina (alinas2000)
  - Collected images of ingredients for beef, salmon, chicken, broccoli, cabbage, carrot, celery, corn, cucumber.
  - Filtered through the images to obtain the most appropriate ones.
- Benjamin (Boweiren)
  - Built the dataset for ingredient recognition, including apple, avocado, banana, lemon, bread, cheese, mushroom, egg, pasta, and rice.
  - Manually checked the dataset and removed the pictures that did not meet the requirements.
- Harrison (hmitgang)
  - Built `data_gathering.ipynb` for cleaning and pickling data from Kaggle's Food.com Recipes and Interactions dataset.
    - Consolidate/simplify ingredients (there were many types of pasta that are now all called "pasta").
    - Filter out recipes that don't contain at least one of our required ingredients.
    - Simplify user recipe ratings and filter out ratings for recipes that have been removed.
- Jane (jwu611)
  - Compiled dataset of ~300 images for each of the ingredients: eggplant, green bean, bell pepper, olive, onion, potato, spinach, tomato, lettuce
  - Filtered out images that were inappropriate for our task, such as images of fully cooked dishes that did not resemble our raw ingredient