

Victoria Vynnychok

Foodie: NLP Food Recommendation Discord Chatbot

Abstract

Foodie is a food recommendation discord chatbot that uses an NLP based model in order to give user's food recommendations. The NLP recommendation model uses Word2Vec vectors as well as cosine similarity in order to recommend foods based on a user specified input of ingredients and keywords.

Problem Description

One universal experience for most of us is deciding what we should eat when we are hungry. In order to make this process easier, I decided to create a model that is able to output food recommendations for users based on their input of ingredients.

Data Description

I use the Food.com dataset, specifically the RAW_recipes csv file in this project which had about 230186 data entries. This data includes columns such as ingredients, description, and ratings. I specifically used the ingredient and name columns for training my model. I also used the description column for name entity recognition, to extract keywords such as "sweet" and to add them to a new data column for a recipe.

https://www.kaggle.com/datasets/shuyangli94/food-com-recipes-and-user-interactions?select=RAW_recipes.csv

Approach

First, I set up a virtual environment to download all needed dependencies for the model and bot. Then, I set up a python Discord bot using documentation. I tested this bot to make sure it connects to a discord server and responses to a user. This was the initial set up before working on the NLP portion.

Next, I found a dataset and created a Google Colab notebook to work in. I loaded the data from my dataset and preprocessed the data. I then tokenized the ingredient data and vectorized it using Word2Vec. I trained a Word2Vec model on the dataset using the ingredient and name columns. Afterwards,

Experiments and Error Analysis

There were many attempts tried at creating both the recommendation model and discord bot.

My first approach involved using SPARQL in order to query wikidata for food related data based on user input. I had trouble in my approach to querying the database for the exact data I wanted which involved all food names and their names and extracting that data. Therefore, after multiple attempts, I decided to find a dataset online from Kaggle instead. I found the Food.com recipe dataset which I determined to be sufficient due to the numerous amounts of tags and ingredients based data for each recipe name.

In relation to the NLP model, the first model did not recommend recipes that included the same ones that the user inputted. After doing some code and error analysis with multiple print statements, I found that the tokens being inputted into the Word2Vec model were characters instead of words that were composed of full ingredients. After removing commas from my data and tokenizing it correctly by word rather than individual characters, I saw large improvements in the accuracy of my model.

Conclusion

The Word2Vec model takes user input and recommends the recipe that is closest to what you want to eat. This model works on a simple basis but can be improved in the future in many ways such as by recommending users the highest rated recipe name, providing links to the recipes, and letting the user ask for food related to the meal time of day such as 'lunch' or 'dinner'. It can also do better to handle user ambiguities by improving to process the data.

References

Discord Documentation to Set up Python Based Bot: <https://discordpy.readthedocs.io/en/latest/>