Ingredients Neural Network

Holly Erickson

DSC 680 Applied Data Science

GitHub Portfolio URL: https://github.com/Holly-E/Cuisine-Project

Which Domain?

What domain is this data going to come from? Please list 10 references (with a brief annotation) to use to make sense of what you're doing with these data.

The domain will be cooking, new recipes, and different cuisines around the world.

References:

1. Bostan, L. (July 17, 2017). *"Ingredient-driven Recipe Generation Using Neural and Distributional Models."* Retrieved from https://lct-master.org/getfile.php?id=2194&n=1&dt=TH&ft=pdf&type=TH

   This is a really interesting project where the input is a list of ingredients and the output is a written recipe that is based on those ingredients.

2. Jain, A., Rakhin, K., Science, G.B., Jodhpur, I.I., India., & Science, C.F. (2015). *"Spices form the basis of food pairing in Indian cuisine."* arXiv: Physics and Society. Retrieved from https://arxiv.org/abs/1502.03815

   Indian cuisine encompasses a number of diverse sub-cuisines separated by geographies, climates and cultures. This is a study in the food pairing in recipes of Indian cuisine to show that, in contrast to positive food pairing reported in some Western cuisines, Indian cuisine has a strong signature of negative food pairing; more the extent of flavor sharing between any two ingredients, lesser their co-occurrence.

3. Fine Dining Lovers, Editorial Staff (Aug 1, 2015). *"Defining 26 Culture's Cuisines by the Ingredients they Use."* Retrieved from https://www.finedininglovers.com/article/defining-26-cultures-cuisines-ingredients-they-use

   An overview of a study done by Priceconomics on distinguishing between different cuisines simply using ingredients. Using a dataset of 26 styles of cuisine from 13,000 recipes posted on Epicurious in 2013, they present their findings.

4. Yau, N. (Sept. 18, 2018). *"Cuisine Ingredients - Looking for the ingredients that make food taste different around the world."* Retrieved from https://flowingdata.com/2018/09/18/cuisine-ingredients/

   This is a study based on the Yummly ingredients dataset. It contains ingredient lists for a bit under 40,000 recipes from 20 cuisines. This amounted to 6,714 ingredients.

5. Brantley, A. (May 29, 2018). *"The Science of Taste: Why Some Foods Taste Good Together."* Retrieved from http://bcbstwelltuned.com/2018/05/29/the-science-of-taste-why-some-foods-taste-good-together/

   Discusses why foods work or don't work well together, based on their compounds. Also, what you need to think about in terms of balance when creating a recipe. Ideally, a balanced dish will incorporate at least three or four of the five tastes — sweet, sour, bitter, salty and umami (savory) — even if it's just a squirt of lemon (sour) or a drizzle of honey (sweet).

6. World Food and Wine. (2005). *"DESCRIBING TASTE AND FLAVOR."* Retrieved from https://world-food-and-wine.com/describing-food

   A helpful resource for teaching the words for describing tastes and flavors of ingredients and dishes.

7. Ahn, Y., Ahnert, S., Bagrow, J. et al. (2011).*"Flavor network and the principles of food pairing."* Sci Rep 1, 196. Retrieved from https://doi.org/10.1038/srep00196

   The presentation on a network-based approach to explore the impact of flavor compounds on ingredient combinations. They created a supplementary dataset that describes ingredient compounds which I will be using in this project.

8. Moyer, M. Tulp, J. (Sept 1, 2013). *" The Flavor Connection. Scientists link common flavor compounds across the world's favorite ingredients."* Retrieved from https://www.scientificamerican.com/article/flavor-connection-taste-map-interactive/

   Includes a flavor map where lines connect foods that have components in common; thick lines mean many components are shared. By comparing the flavor network with various recipe databases, the researchers conclude that chefs do tend to pair ingredients with shared flavor compounds—but only in Western cuisine. Dishes from a database of recipes from East Asia tend to combine ingredients with few overlapping flavors.

9. Trekhleb, O. (2020). *"Generating cooking recipes using TensorFlow and LSTM Recurrent Neural Network: A step-by-step guide."* Retrieved from https://www.kdnuggets.com/2020/07/generating-cooking-recipes-using-tensorflow.html

   A character-level LSTM (Long short-term memory) RNN (Recurrent Neural Network) is trained on ~100k recipes dataset using TensorFlow. The focus is not on flavors, but on generating a readable recipe that could be made in real life.

10. Shane, J. (2016). *"The Silicon Gourmet: training a neural network to generate cooking recipes."* Retrieved from https://aiweirdness.com/post/140219420017/the-silicon-gourmet-training-a-neural-network-to

    Similar to resource #9, this is the overview of the process for creating a recipe generator using recipes downloaded from a collection by David Shields, via a neural network.

Which Data?

What is the dataset you'll be examining? Please provide a codebook if there is one or a link to the dataset as well as a detailed description.

I am hoping to combine three datasets:

1. The dataset used in the Kaggle Competition "What's Cooking?". I have worked with the data before in that competition, however I lost that code, and I will be achieving a different, more complex mission with this model than the goal of that competition.

   Each row in the dataset includes the recipe id, the type of cuisine, and the list of ingredients of each recipe (of variable length). The data is stored in JSON format.

2. Dataset containing ingredients and their flavor compounds. Source: Ahn, Y., Ahnert, S., Bagrow, J. et al. (2011).*"Flavor network and the principles of food pairing."* Sci Rep 1, 196.

   Retrieved from https://doi.org/10.1038/srep00196

   Supporting online material: https://static-content.springer.com/esm/art%3A10.1038%2Fsrep00196/MediaObjects/41598_2011_BFsrep00196_MOESM1_ESM.pdf

3. Retrieved from an API called The Meal DB. https://www.themealdb.com/

   This site allowed users to easily access the data for free if it is for educational purposes. I created a request that would pull random recipes from their website using JSON. I will be using this data in my neural network, but not in the cuisine based classification algorithm.

   Each row includes: Recipe name, time to make, number of ingredients, ingredients list, number of steps, list of steps, keywords, description and nutrition information.

Research Questions? Benefits? Why analyze these data?

How are you proposing to analyze this dataset?

The idea behind this model is for new dish creation inspiration. The input to the model is one ingredient that you would like to cook with, the output is a list of ingredients it could be used with, and the top cuisines this combination of ingredients would most likely be found in.

The neural network will hopefully learn which combinations of ingredients work well together based on their flavor compounds, and recommend unusual and interesting combinations that chefs or home cooks may like to try.

What Method?

What methods will you be using? What will those methods provide in terms of analysis? How is this useful?

I will be using a neural network to offer the chef a set of ingredients to try together. The neural network will need to learn the flavor compound of ingredients, and will be trained on existing recipe ingredient combinations.

I will also use a classification algorithm that classifies a list of ingredients into the type of cuisine they belong two. The output from the neural network would be the input to this algorithm. That way the final model will be able to offer a list of ingredient pairings based on the initial input ingredient, as well as the type of cuisine this flavor combination most likely would be found in.

Potential Issues?

What challenges do you anticipate having? What could cause this project to go off schedule?

I don't have much practical experience working with neural networks. I have coded one from scratch in a course on them, but I have not used libraries. I am excited to have the opportunity to give it a shot.

Ideally I will be including a flavor profile for the ingredients as data for my model. It will be interesting to try to combine several datasets that have not been designed to be combined, and hopefully that will go smoothly.

I have an idea in my mind of what I'd like the model to output, but I am not sure exactly how to make it happen. It will hopefully be a fun challenge.

Concluding Remarks

Tie it all together. Think of this section as your final report's abstract.

Imagine you are walking through the grocery store or farmer's market in search of inspiration. A new ingredient catches your eye and you decide to buy it. What ingredients might you pair with this?

The ingredient is the first input to a neural network and classification model. The second input is the number of ingredients you would like to include in the dish. The model output is a list of recommended ingredients to pair with. It also provides the type of cuisine this flavor combination most likely would be found in, in case you would like to do some research.