Machine Learning in Network Science

So…what do you wanna eat?

Project proposal

| Aiza AVILA  DSBA M2  ESSEC Business School and Centrale Supélec, France  aiza.avila-canibe@student-cs.fr | Aline HELBURG  DSBA M2  ESSEC Business School and Centrale Supélec, France  aline.helburg@student-cs.fr | Joshua Jehau FAN  DSBA M2  ESSEC Business School and Centrale Supélec, France  joshua.fan@student-cs.fr | Xiaoyan HONG  DSBA M2  ESSEC Business School and Centrale Supélec, France  xiaoyan.hong@student-cs.fr |
| --- | --- | --- | --- |

**ABSTRACT**

In this project, we propose a machine learning-based approach to explore the similarity between the ingredients of different cuisines and determine their pairing of flavors. We construct ingredient networks for each cuisine by representing ingredients as nodes and co-occurrences as edges. We employ machine learning’s algorithm to model the relationships between the ingredients, and use it to generate feature importance scores for each ingredient. By analyzing these scores, we identify the key flavor components of each cuisine, and compare them to find similarities between different cuisines. We then use this information to recommend dishes based on their similarity in terms of flavor pairing. We assume that our approach will be effective in identifying flavor similarities between cuisines, and can be used to recommend dishes that are likely to be enjoyed by individuals based on their culinary preferences. This approach has the potential to enhance the culinary experience for individuals by providing tailored dish recommendations based on their flavor preferences.

CCS CONCEPTS

• Network Science • Machine Learning   • Graphical Network

KEYWORDS

Food Computing, Food Pairing, Flavor, Ingredient Network, Similarity

**INTRODUCTION**

Machine learning and network science have both emerged as important fields in recent years, and their combination has led to new and exciting applications. The ability of machine learning algorithms to learn patterns and relationships from large datasets, combined with the power of network analysis to model complex systems, has led to a surge of interest in this interdisciplinary area of research. Machine learning techniques have been used in network science to analyze large-scale networks and uncover hidden patterns and structures. Their application scope is broad - from bioscience to social media and customers’ recommendations. The ability to analyze complex systems and uncover hidden patterns has already led to important applications in fields such as healthcare, social media, and culinary arts. With continued research and development, this interdisciplinary field is sure to lead to even more exciting breakthroughs in the years to come. With this project, we aim to dig into one application that is of high interest to us, with a high significance and technically challenging.

**1 PROBLEM DEFINITION**

As a result of our participation in the master's program, each of the students has been able to exchange points of view on different topics with our peers. And perhaps the most repeated has been *food*. Some of us miss the dishes of our homeland and have had to adapt to the ingredients available in France, making the culinary experience sometimes not 100% authentic. Having a shared feeling of longing, we have recreated regional dishes on multiple occasions, sharing experiences and flavors with our colleagues, and on more than one occasion we have been surprised by the similarities and differences we have in the use of ingredients. Motivated by this fact, we see in the analysis of ingredients a possibility to break down the flavors that unite us and those that make us unique.

We propose to build a *dish recommender* based on the similar flavors that exist in other regions, in order to get closer to other similar cultures. However, this project may have other purposes, such as a recipe suggestion engine or dish recommendation engines, such as those offered at <https://www.foodpairing.com/>

**2 METHODOLOGY**

Our project is based on data from YY Ahn's work (see reference [1]), including 1530 food ingredients, flavor compounds and 36,781 edges. We also have a database of 7,000 recipes.

To complete the challenge and build our dish recommender we plan to follow the next steps:

1. Disaggregate the recipe database to have a list of ingredients. This step involves breaking down each recipe in the database into a list of its constituent ingredients. For example, if a recipe calls for "pasta with tomato sauce and basil", this step would break it down into three ingredients: pasta, tomato sauce, and basil.
2. Perform weights of ingredients depending on the recipe in which they are. This step involves assigning weights to each ingredient based on its quantity in the recipe. For example, if a recipe calls for 2 cups of pasta and 1 cup of tomato sauce, pasta would have a weight of 2 and tomato sauce would have a weight of 1.
3. Map ingredients and recipes. This will allow us to easily identify which ingredients are used in which recipes and have an overall insight.
4. Relate ingredients with "flavors". This step involves grouping similar ingredients together based on their flavor profile. For example, grouping paprika, pepper, chili into "spices", to mention some examples
5. Apply a classifier (random forest classifier for example), in order to classify regional dishes
6. Apply a measure of similarity (for example jaccard) to recommend those dishes that are close to the desired flavor. For example, if we choose "spicy pasta" it will return dishes that are more similar to the search
7. Finally, we will use a visualization package available on Python such as NetworkX or PyVis. We will build a graph network, with each node representing an ingredient and the edge between nodes representing the number of flavor compounds they are connected to. Given the data used, we are expecting to have a really dense and unreadable graph therefore we are planning to only keep the ingredients with significant connections to others.

In previous works, the main focus was on finding the similarity between the ingredients of different cuisines, in order to find the pairing of flavors. We want to push this approach towards dish recommendations based on the similarity between them.

On Python, we are planning to use the following libraries : pandas, sklearn, networkx, numpy, scipy, nltk, matplotlib, seaborn. Others may be added depending on our needs.

**3 EVALUATION**

To evaluate the performance of the model we propose to define a ranking of the dishes (using Discounted Cumulative Gain) in order to ensure that the model shows the top relevant dishes.

The relevance of each item will be represented by a relevance score. The position of each item is represented by a discount factor that decreases as the item's position in the list increases. The discount factor is often calculated using a logarithmic function.

The formula for calculating DCG is:

Where rel(i) is the relevance score of the item at position i in the list, and n is the length of the list.

We are also planning to perform some user tests, generate multiple queries and perform manual research on the dishes and the ingredients, to give us an idea if indeed the similarity makes sense, or if it doesn't.

**REFERENCES**

[1] Ahn, YY., Ahnert, S., Bagrow, J. et al. Flavor network and the principles of food pairing. Sci Rep 1, 196 (2011). https://doi.org/10.1038/srep00196

[2] Data repository https://yongyeol.com/

[3] Maija KĀLE, Ebenezer AGBOZO. Utility of Large-Scale Recipe Data in Food Computing https://elar.urfu.ru/bitstream/10995/111694/1/2-s2.0-85109959710.pdf

[4] Weiqing Min, Shuqiang Jiang, Linhu Liu, Yong Rui, and Ramesh Jain. 2019. A Survey on Food Computing. ACM Comput. Surv. 1, 1 (July 2019), 42 pages. https://arxiv.org/pdf/1808.07202.pdf

]5] Chun-Yuen Teng, Yu-Ru Lin, Lada A. Adamic (May 2012) Recipe recommendation using ingredient networks https://arxiv.org/pdf/1111.3919.pdf