Maël Wildi
Karen Preitner
Raphael Strebel
Elias Gajo

Strategic Recipes

NTDS
December 1, 2019

Motivation

Our motivation is to limit food waste. Food waste remains an important problem in our society. Our approach to improve this situation is to reduce the waste on the consumer side. One way to tackle this would be to propose recipes that maximize the number of ingredients that people already have at home, while minimizing the amount of new ingredients they have to buy.

Data

We plan to use the dataset `Cooking recipes`. It contains the ingredients, the rating of each recipe, along with the number of reviews and the cooking time. We start by scraping a few of the 163 websites available in the dataset, enough to keep an appropriate amount of recipes. We then store the dataset into a pickle file, in order to avoid computing the scraping part each time we open the notebook.

Limitations

Although we have access to the quantity of each ingredient in the recipe, it might be hard to convert them to a single metric (cup, spoons, grams, liters,...). For this reason, and as this is not the goal of the course, we will limit ourselves to sets of ingredients without considering the weight/quantity of each of them inside recipes.

Graph

As we deal with sets of ingredients, using the Jaccard distance as edge weighting method seems the most appropriate. This allows us to build a weighted graph, where a node represents a recipe and an edge measures the Jaccard similarity between two recipes.

Goals

Our main goal is to provide a user with a highly rated recipe, that contains as many ingredients available as possible. More concretely a user can input a set of ingredients, which we would consider as a new node, and we output the closest well rated recipes in the graph.

Should there be no such close recipe, we could output one that requires the least amount of new ingredients. One way to optimize this goal would be to start by forming clusters of recipes (e.g. sweets, starters, cold dishes,...). That way, a new input node doesn't have to compute its similarity towards all existing nodes.

References

Input fridge image : https://tse2.mm.bing.net/th?id=OIP.XuqTiL8pKakhcRTLfkgGKwHaGD
Output fridge image : https://www.masterfile.com/image/en/700-00603409/leftovers-in-fridge
Cooking recipes dataset : http://infolab.stanford.edu/~west1/from-cookies-to-cooks/recipePages.zip