Recipe for Success:

Optimizing meals under dietary constraints

Category: General Machine Learning

Members: James Ordner (jordner), Benjamin Share (benshare), Zack Cinquini (icinquin)

Motivation

When faced with thousands of search results, finding a recipe online that aligns with your personal preferences and potential dietary restrictions can be an intimidating task. It can be difficult to pinpoint which aspects of a dish are crucial to its success, and which are more flexible. We plan to design a machine learning model that will create a mapping of recipes (as captured by their constituent ingredients) to success as measured by online reviews from people who have made and tasted the meals. Then, we plan to develop an algorithm that, given a set of dietary constraints, generates successful recipes under those parameters by adjusting ingredient ratios and detecting logical ingredient substitutions.

Methods

Our general strategy will be to treat quantities of ingredients as features and recipe ratings as labels. Baseline models should include linear regression (identifying "good" vs. "bad" ingredients) and weighted averages over ingredients within a dish type. Recipes receive weighting based on their rating "deviation" from the average; their use of ingredients can thus inform our notion of an ideal recipe. We also hope to use Naive Bayes as a form of classifier for "buckets" of recipe ratings (divided by stars or partial-stars) and should serve as a reality check for the accuracy of other methods.

Ultimately, we hope to develop several more sophisticated approaches as well. Using vector embeddings, we aim to extract relevant features from ingredients and build a system for semantic ingredient parsing. One important application of these embeddings will hopefully be the ability to recommend substitutions based on spatial similarity (i.e. vectorized nearest-neighbors). If possible, we'd also like to implement a deep learning approach to the problem, consisting of a multilayer neural network. By allowing the network to learn a more sophisticated set of interactions between inputs, it should be able to better understand the tradeoffs between ingredient inclusion. One relevant test case will be to see if a network can be trained on broader categories than the simple models, implying that it understands the existence of discrete subtypes within the category.

Intended Experiments

We will test our model on a few specific foods, such as lasagna and chocolate chip cookies. Then, we will will optimize recipes subject to various dietary constraints (less sugar, low sodium, no dairy, etc.).

To evaluate the accuracy of our model's recipe ratings, we will train on a subset of the data and test our model's predictions on the remaining portion.

Testing the constrained optimization part of our project will be less straightforward, as it is difficult to test the quality of a novel recipe without cooking and tasting it. We plan to evaluate these results qualitatively, by checking that no unusual quantities of ingredients were used and that the recipe includes the essential types of ingredients for a specific recipe (for example, a cookie recipe cannot contain all dry ingredients like flour or sugar).

Prior Research

Recipe Recommendation Using Ingredient Networks: Chun-Yuen Teng, Yu-Ru Lin, and Lada A. Adamic, 2012

Dataset

Allrecipes.com