Recipe for Success

Optimizing meals under dietary constraints

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

Introduction

Finding a healthy recipe online that aligns with your personal preferences and dietary restrictions can be a daunting task. Using data from allrecipes.com, we constructed machine learning models that map recipes (as captured by their constituent ingredients) to success as measured by online ratings. Using vector representations of our ingredients, we develop a methodology for detecting logical ingredient substitutions.

Dataset and Features

Lasagna: 259

Brownies: 383

Cookies: 4703

"Banana Nut Brownies"

- Rating: 4.39
- Reviews: 1322
- Servings: 20
- Ingredients: [2 cups white sugar, 1 cup butter, 1 1/2 cups all-purpose flour, ... 1 ripe banana, mashed]

Servings: 20

["2 cups white sugar",
"1 cup butter",
"1 1/2 cups all-purpose flour",
...
"1 ripe banana, mashed"]

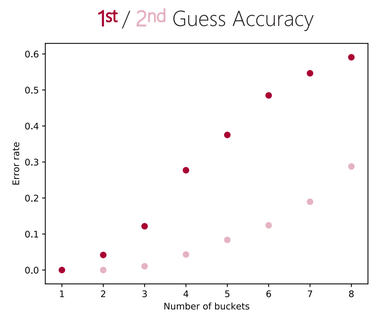
- [[0.67, sugar"],
- [0.33, "butter"],
- [0.5, "flour"],
- ...
- [0.83, "banana"]

Recipe Rating Prediction

Naive Bayes

Model: We apply the Naive Bayes multinomial model with Laplace smoothing. A flexible bucketing schema discretizes the continuous ranges of ratings from one to five stars.

Data and Features: We filter the recipes to include only ingredients that appear in at least 10 meals. Recipes are randomly partitioned into training (80%) and test (20%) sets.



Best Cookie Ingredients

- "cream cheese" (3.39)
- "peanut butter" (3.47)
- "semi-sweet chocolate" (3.51)
- "marshmallow" (3.54)
- "peanut butter cup" (3.76)

Worst Cookie Ingredients

- "coconut sugar" (1.92)
- "peppermint extract" (1.94)
- "anise oil" (1.95)
- "orange extract" (1.95)
- "almond milk" (1.95)

Sample Generated 5-star Recipe

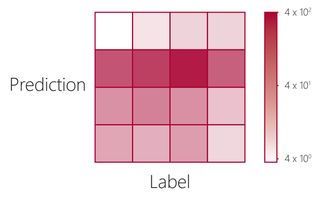
2 tbsp vegetable oil
2 tbsp condensed milk
2 tbsp marshmallow
2 tbsp brown sugar
2 tbsp walnut
2 tbsp shortening
1 1/3 cups all-purpose flour
1/2 cup butter
1 floz vanilla extract
3/4 cups of white sugar
2 tbsp almond
4 tbsp of chocolate cake mix
2 tbsp cocoa powder
2 tbsp raisin
2 tbsp pumpkin
2 tbsp confectioners' sugar
2 tbsp water
3/4 cups semi-sweet chocolate
3 tbsp of peanut butter cup
3 eggs

Rating Neural Network

Model: We implemented two neural networks—one for classification, and one for regression. Both use one hidden layer and sigmoid activation.

Data and Features: Again, we filter the recipes to include to only track ingredients that appear in at least 10 recipes. Recipes are randomly partitioned into training (70%) and test (30%) sets.

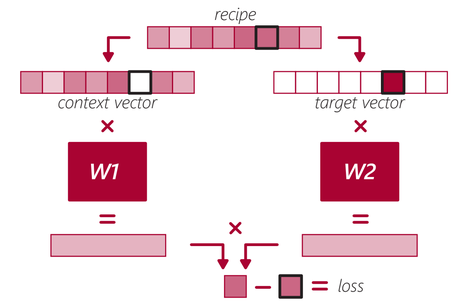
Classification Accuracy: 53%



Ingredient Substitution

Word2vec Neural Network

Model: We adapted Mikolov *et al.*'s word2vec model for generating vector representations of features, allowing us to synthesize ingredients as vectors in a high-dimensional space.



Substitutes for "chocolate cake mix"

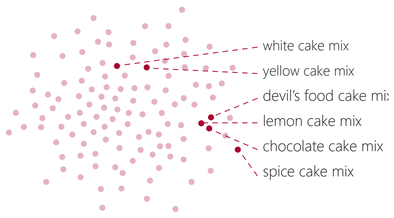
Best Ingredients

- "devil's food cake mix"
- "lemon cake mix"
- "coconut"
- "chocolate pudding mix"
- "marshmallow"

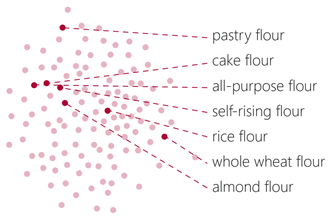
Worst Ingredients

- "semi-sweet chocolate"
- "white sugar"
- "brown sugar"
- "peanut butter cup"
- "all-purpose flour"

Embeddings of "cake Mix"



Embeddings of "flour"



Future Directions

Future work would involve integrating our models for rating prediction and ingredient substitution into one tool to generate highly-rated recipes given a set of dietary constraints. Doing this would likely require more data for many types of meals. There are a number of more nuanced approaches that can be further explored, such as reverse-engineering our rating network. The tools we've explored should make this process fairly straightforward.