

Capstone project: Predicting success of food and nutrient combinations

Mike Touse

October 30, 2016

The goal of this project is to predict the success of foods for high-endurance activities such as extremely long-distance hiking (as in Appalachian or Pacific Crest Trail through-hikes). While the expected metric will be sustained energy per weight, other measures of performance may include sentiment analysis, reported performance impact, or comparison with known long-term, high-performance nutritional requirements. While the project focuses on hiking, it could potentially extend to other activities such as school-lunch programs or similar public, nutrition-focused programs.

1 Problem to be solved

The multi-billion dollar performance food industry includes countless options for consumers. Because the industry supports nearly every endurance and high-performance activity, most products are generically marketed for features such as *energy*, *endurance*, or *recovery* with no effort to optimize for the specific requirements of individual activities.

One activity in particular, long-distance backpacking, carries with it a particularly challenging requirement: nutrition-to-weight ratio. While the requirement to maximize energy:weight is challenging in itself, the problem becomes even more difficult when the activity duration is measured in months instead of hours or days, as is needed for treks along the Appalachian and Pacific Crest Trails. In activities lasting so long, simply selecting the highest caloric density food is not adequate due to the body's requirement for a variety of nutrients, as well as the tendency for hikers to simply get tired or bored with a single food.

Long-distance treks call for weeks or months of menu and logistics planning. The initial problem was simply posed as a question of which foods are best to take backpacking, and whether particular combinations of foods, ingredients, or nutrients have undiscovered benefits or efficiencies.

2 Client description and actions

As previously stated, there are several potential clients for this analysis:

- Hikers preparing for long-distance through-hikes can use the study to determine a number of foods that optimize nutrients and energy for the given weight.
- Companies supporting high-performance activities can use the analysis to predict the potential success of new or proposed products.
- While not projected for this particular project, the analysis could be extended to optimize for other features (besides weight) such as cost or dietary restrictions.

3 Data to be used

Several data sources have been reviewed or postulated including:

- The USDA Nutrition Database provides nutrition data for many foods, and could be coupled with recipes or ingredient lists from energy foods to predict the most useful or successful products or nutrient combinations.
- world.openfoodfacts.org provides an open database that is compiled through user contribution to provide nutrition data for products (>100,000 currently). It includes data similar to the USDA Nutrition Database referenced above.
- Product providers (websites) such as REI.com, Amazon.com, or others can likely provide product reviews for the products detailed on openfoodfacts or USDA databases.
- Recipe websites such as spoonacular.com/food-api, developer.edamam.com/edamam-recipe-api, and food2fork.com/api allow recipe searches.
- Hiking blogs spend a significant amount of time discussing food.

4 Approach

The overall approach, to be further developed in the next stage of the project, is essentially to find foods that are rated highly for hiking, and cross-reference those foods to nutrition content. The supervised learning will then be used for prediction of combinations of features (nutrients/ingredients) versus weight of as many foods as possible with the successfully mined datasets.

5 Deliverables

Delivery of this capstone project would be through ipynb on github and a presentation.