Food Categorization

Prepared by
Alshouiliy, Khaldoon M06967001
Park, Kyungmook (Chris) M07068980
20CS6037 – Fall2014
Due: 11:59PM 11/15/2014

University of Cicinnati
Collage of Engineering and Applied Science

Abstract

In this project, we present a method for categorizing various foods into known food groups by using supervised learning. A food with details such as name, nutritional information, method of cooking, serving size, and description can be classified into a particular food group. In order to accomplish this, we used the food data prepared and provided by the United States Department of Agriculture (USDA) to build features for such a classifier. We will implement our method to 90% of the USDA food data to train the data. Then we will use the rest 10% of the data to test and predict the food group a food belongs in based on particular criteria. We hypothesize that we can achieve 95% accuracy of the food group assignment with the proper constraints imposed on the USDA food data. Our implementation will employ multiple Machine Learning techniques and compare their results to one another. We will use the following Machine Learning techniques: neural network, genetic algorithm, and naïve-Bayesian.

Introduction

Machine Learning has been getting an increasing attention in the art of categorization. Many investors and companies are interested in intelligently categorizing various itineraries obtained from Big Data. Such categorization efforts help in deciphering the meaning of all the data collected. It is predicted that making data-driven decisions will become even more important as technology progresses.

We will be using the USDA food data that has a plethora of different categorization groups. For example, Kale belongs in 9 different categories, raw, frozen and unprepared, cooked and boiled, cooked and drained without salt, and so on. Of course, additional categorization can be made based on the food's characteristics. These characteristics include nutritional contents, which food group the food belongs in, which food sub group it belongs in, and what kind of food is similar to it. The data file we will be using has an Excel, comma delimited format: Food_Data.xslx.

On the other hand, we will implement our method by first training the 90% of the USDA food data. Then, we will use the other 10% of the data to test and predict the food group a food belongs in based on particular criteria as in Figure 1.

In this case, our implementation will employ multiple Machine Learning techniques and compare their results to one another. We will use the following Machine Learning techniques: neural network, genetic algorithm, and naïve-Bayesian. As a result of this project, we expect that we can achieve over 95% accuracy of the food group assignment with the proper constraints imposed on the USDA food data.

Project Methodology Diagram

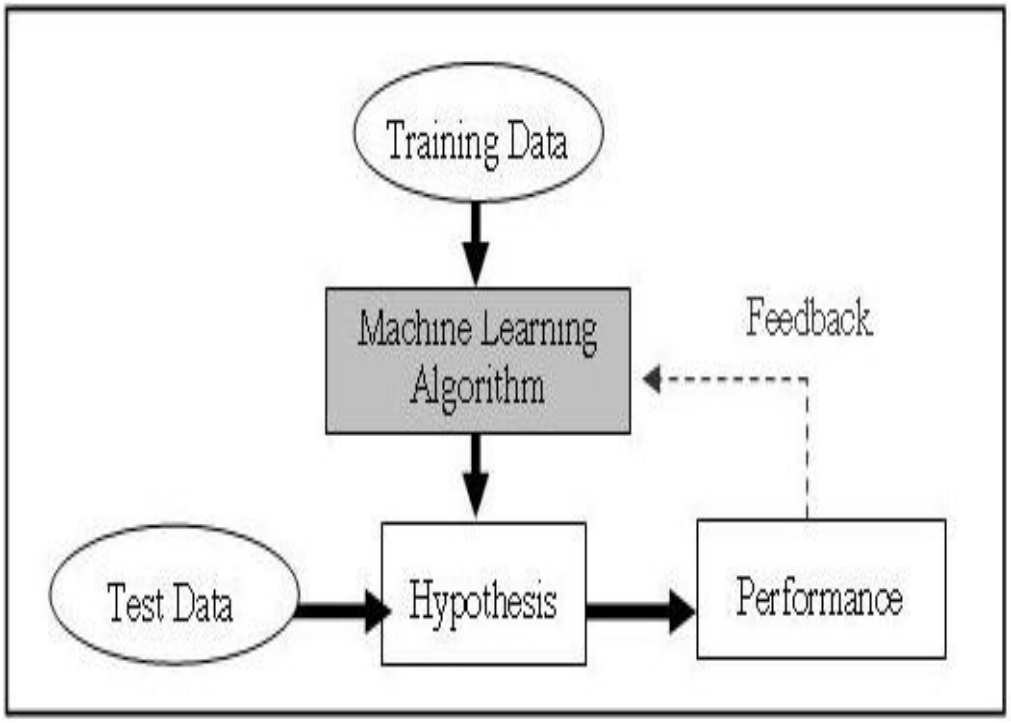

Figure 1 - Method Diagram

Problem Statement

By categorizing and clarifying the nutritional information, many people will be able to select their food based on their dietary needs. This is crucial since proper diet has tremendous impact on one's health. The consequence of improper diet includes obesity, diabetes, and hypertension. The effort mentioned in this proposal will help people to choose appropriate meal in order prevent health issues.

Project Objectives:

1. Provide a clarifying food categorization so that people can choose the best meal for themselves by implementing multi-algorithm on the USDA food data.
2. Consider which one of the Machine Learning algorithms is suitable for people to choose their health food.
3. Achieve 95% accuracy of the food group assignment with the proper constraints imposed on the USDA food data.

Conclusion

We will implement various Machine Learning algorithms to categorize food into a proper food group. The categorization will depend on the nutritional information such as the amount of protein, carbohydrates, and fat. Furthermore, we will be incorporating the food groups and food sub groups already present in the USDA dataset in order to produce the most accurate predictions. In order to achieve this accuracy, we need to have a functional knowledge in nutrition and practical knowledge for implementation of all the algorithms we are going to implement, i.e., Neural Network, Genetic Algorithm, and Naïve-Bayesian.