Restaurant Successes and Failures

Team: Jason Feng, Naho Kitade, and Matt Ritter

Problem Statement

The restaurant industry has an incredibly high turnover rate. Many new restaurants do not survive their first few years in business. In addition, the restaurant business is flooded with various factors that impact success (location, timing, atmosphere, food quality, etc), such that it can be extremely difficult to create a successful restaurant. We hope to analyze Yelp data to determine if there are certain factors that are most important for a successful restaurant. This study will be conducted individually for different cities, since geography plays an overwhelming role in food culture. In theory, our studies will reveal certain key points of success for a restaurant business.

We will define our inputs as the following pieces of Yelp data:
* See Table of Features on second page

We will define the following formula measuring restaurant success for our output:

- Let R be the number of reviews the restaurant will receive (avg of 53 reviews in data set)
- Let S be the number of stars the restaurant will receive (avg of 3.46 reviews in data set)
- Let 70% of the output come from the number of stars. Since there can be between 1 and 5 stars, each star is worth 14 percent.
- Let 30% of the output come from the number of reviews. Since the average number of reviews is 53, let restaurants with average score receive 15 out of 30 points available.

$$14S + r(R) = SUCCESS\ SCORE\ (range: 0{\sim}100)$$

Where $r(x)$ is defined as:

$$r(x) = \frac{120}{(1\ +\ 9e^{-0.02\ (x+100)})^{0.4}} - 80$$

A logarithmic function with upper asymptote at 30.

Data Sets

We plan to use existing Yelp data for our study. Yelp releases a large dataset for competitions and other academic uses. The page from which we requested the data can be found here.

This data contains information on approximately 40,000 businesses clustered around five geographic areas. Of these, our analysis has shown that 14,303 of these are in the "Restaurants" category. In addition, the data contains 1,125,458 reviews on the businesses in the data, 252,898 users, and 403,210 tips. A tip is a helpful piece of information about a business, submitted by a user. The data is clustered around a few geographic areas. This will allow us to compare our results across different locations, since food culture can vary wildly depending on location.

We will be able to use some of this data (70 - 80%) as our training data and hold out a portion of the data (10 - 20%) as test data.

Refer to Table of Features on page 2 for detailed description on features and their types.

Suitable Methods

We have posed our problem as a classification problem, and the Yelp dataset allows for supervised learning. We plan to implement the feedforward Artificial Neural Network (ANN) model with multilayer perceptrons and a backpropagation training algorithm. We will use sigmoid neurons rather than simple perceptrons in our model as to deal with non-binary (continuous from 0 - 1) output needs. Similar analyses of the restaurant industry have been conducted in past using ANN models as well. [3]

Reference for general detailed method description – [1] [2]
Chapter 1 and 2 of book *Neural Networks and Deep Learning*

Milestone Goals

By the milestone, we will have implemented a rudimentary neural network to test our problem statement. Fortunately, we already have access to our dataset, so we will not need time to collect the data. We will also have randomly sampled 70% of our data to be used as test data and the remaining 30% to be used as training data.

Table of Features

Table of Features	
Inputs	Type
Take out	Boolean
Good for breakfast	Boolean
Good for lunch	Boolean
Good for dinners	Boolean
Caters	Boolean
Noise level	{quiet, average, loud, very loud}
Takes Reservations	Boolean
Delivery	Boolean
Parking	Boolean
Has TV	Boolean
Outdoor Seating	Boolean
Attire	{casual, dressy, formal}
Alcohol	Boolean
Waiter Service	Boolean
Accepts Credit Cards	Boolean
Good for Kids	Boolean
Good for Groups	Boolean
Price Range	float
Latitude	float
Longitude	float
Additional features acting as tags on Training Data	
Star Rating	float (0 up to 5)
Number of reviews	int

References:

1. Nielsen, M. (2015). Neural Networks and Deep Learning. Determination Press. [online] Available at: http://neuralnetworksanddeeplearning.com/chap1.html [Accessed 21 Jan. 2015].
2. Nielsen, M. (2015). Neural Networks and Deep Learning. Determination Press. [online] Available at: http://neuralnetworksanddeeplearning.com/chap2.html [Accessed 21 Jan. 2015].
3. Youn, H. and Gu, Z. (2015). JSTOR: Tourism and Hospitality Research, Vol. 10, No. 3 (JULY 2010), pp. 171-187. [online] Jstor.org. Available at: http://www.jstor.org/stable/pdfplus/23745462.pdf?acceptTC=true [Accessed 20 Jan. 2015].