Word Pronunciation and Travel Destination Appeal

Faris Sbahi, Nisarg Dabhi, Alex Salas, Jonathan Rub

It is important to create a narration about your travels, which starts at correctly pronouncing your destinations. We wanted to know whether or not a traveler can pronounce a city destination affected their propensity to travel there. We decided to focus just on American travelers making trips out of the country. We encoded difficulty rankings for every city destination outside of the United States. To do this we looked through the 5,000 most popular words and parsed them into the most popular syllables. From the most common english syllables we were able to determine a ranking to show which City names were the easiest and hardest to pronounce. Running the numbers associated with the difficulty through a random forest algorithm eventually yielded results showing the importance of easily pronounceable name for tourism.

**PCA Model Plot**

The PCA graphs showed that the variance by destination ID is explained mostly by one Principal Component of the graph. We used 5 Principal Components to model the data, which captured most of the destination data variance while also not having an overwhelming number of principal components.

**Network**

It is natural to visualize the effect of simplicity in destination name by using a network. We created a directed graph that shows the original destination (U.S. travelers) and the international destination and furthermore averaged the difficulty of the cities in that international destination which is communicated by darkness in color.

**Random Forest Algorithm Generated Decision Tree**

Random forests extend decision tree models by constructing a multitude of trees (indeed, a forest) at training time. Outputted is the class that is the mean regression of the individual trees.  
Hence, random forests are particularly practical in comparison to decision trees due to the emergent and inherent property that the model will correct for the trees’ habit of overfitting to the training set. Indeed, a decision tree is entirely formed from the training data set resulting in large generalization errors, taking a sample of the data and comparing its performance to the rest of the training set (out of bag).