**Inferring Restaurant Characteristic Preference by Geographical Location using Yelp Reviews**

Bogdan Kotzev, David Slayback, Xinhai Xu, Tyler Budisky

{btk15, dts24, xix35, tsb20}@pitt.edu

**Introduction**

* One of the most important decisions when starting a business is choosing the correct location.
* This choice often proves critical in the overall success of a business. This is especially true with restaurants.
* One important factor in this location decision is customer preference for a given area.
* With the rise in popularity of various review sites and e-commerce platforms, this useful information is readily available and in abundance. One example of this would be Yelp, specifically Yelp restaurant review data
* In this project, we intend to design a system that uses natural language processing (NLP) techniques to analyze the restaurant review data and then characterize restaurant preference per geographical location based on the analysis.
* Our system will first preprocess the reviews, searching for key characteristic phrases.
* Then, our system will extract the key descriptive features from these reviews, storing the most common features exhibited by restaurants in a given geographical location.
* The resulting regional characterization would help increase the chances of selecting a good or even the best location to build a successful restaurant business.

**Motivation**

* Many different cities or regions are well-known for various types of cuisine. Often, this is a byproduct of the local area customer having preference to that type of cuisine.
* Further, customers in certain areas may prefer restaurants that exhibit certain characteristics over others based on location. For example, customers in a downtown setting would more than likely prefer restaurants with fast service times while customers in a more rural setting may prefer restaurants with a more scenic view.
* The same is often true with different parts of the same city or region. For example, average customer preference in the northern area of a given city is often different than that of the southern area of the same city.
* The Yelp data contains a multitude of restaurant reviews, many of which contain very useful information for this project.
* However, not all of these reviews contain helpful or useful information. Some reviews may be too opinionated or too vague to help characterize a given restaurant, which is why we will focus more on those reviews that are voted useful by other users.
* **Describe how system uses useful review files in more detail (much like extracting aspect-feature pairs as described in example Zhang proposal)…**
* This information of knowing the propensities of customers in a given area would be invaluable for potential restaurant purchasers and developers.

**Approach**

* **(very similar to Zhang proposal)**
* Preprocessing: remove “noise” words/phrases from reviews (e.g. “Haha”, “lol”, all caps, multiple exclamation points, etc.)
* Key feature Normalization: consolidate similar features (e.g. “awesome”, “great”, “amazing”, “fantastic” = all have same idea (good))
* Phrase Detection? (I like this idea from the Zhang proposal)
* Key Feature Selection Model
  + Need to define metrics for calculation
  + E.g. Given city or region, need to define which restaurants to look at
    - Maybe also have user give restaurant category along with region to get more specific results?

**Evaluation**

* User study?

**Milestone**

* Already listed on course website, don’t think we need this section

**Responsibility**

* Also not sure if too early to define this…