Joshua Pham (jhp2352), Dune Blum (djb3657)

LIN 350: Computational Semantics

Katrin Erk

9 November 2016

Intermediate Report

**Research Questions: Any Changes?**

We noted the feedback that our initial questions were not really research questions and decided to rework them. We will still be looking at feature-based sentiment analysis on reviews. We have, however, decided to create some new research questions based off our same method. Our first question is to see how accurately a model can derive topic keywords that describe each feature we examine. Our work has us using a dataset of reviews that has been pre-annotated according to the feature they most describe. For example, with our dataset of hotels, these features may be things like room cleanliness and service. These labels allow us to explore aspects of the reviews that best answer our research question. Various aspects, from keywords, to adjectives, to noun phrases and other grammatical structures, may have better discerning power than others, due to how they vary in abundance in our dataset. Aspects with greater abundance will allow us to better answer our question. We expect that identifying these aspects will scale to larger real-world datasets, such as online forums instead of a pre-assembled collection of reviews. If we have time, we will examine the performance of our model on online forums. This kind of practical use is more along the lines of our original motivation: to automatically categorize forum discussion, which is chronologically ordered, so that the opinions are easier to study. Should we correctly categorize the data and identify the features noun phrases refer to, we may also choose to analyze the polarity of the adjectives included in each case in order to essentially model the collective reactions to each feature.

**Method: Any Changes?**

Our method of parsing remained largely the same. So far, we have used BeautifulSoup to parse XML/HTML files for tags and their corresponding text. From an XML file containing polarity and category of a review, as well as the review sentence, we obtain structured Python objects. We have studied the sentences themselves in several ways, after, of course, removing all punctuation. One method we have experimented with is extracting all adjectives using NLTK, and creating mappings from categories to the adjectives used within each category’s sentences. While we do bin the adjectives into each category, we also bin the categories into each adjective. Studying the category diversity here is a way to gain insight on the discerning power of various adjectives. Another method we have tried is to extract word sequences that match grammatical structures from the review texts. We use NLTK’s RegexpParser to define patterns to look for. We then analyze the matches as a whole or examine the nouns or adjectives in these matches. Our intent is to use these results in a similar fashion to our approach with only adjectives – that is, to study discerning power from term-frequency-inverse-document-frequency.

**Describe the data that was obtained: source, size, anything else that was relevant**

We obtained most of our current data in which we plan to train our model on through open-source projects. We have 266 sentences of hotel reviews, 685 sentences of restaurant reviews, and 761 sentences of laptop reviews. After we build our model from this set of data, we hope to be able to easily parse data from forums and other similar review sites and be able to analyze that, but the purpose of our current set of data is simply to largely train our model. We feel pretty comfortable with the data we obtained, as well as the data we are capable of obtaining, as even if we have trouble parsing forums, there is a multitude of data available through open-source projects which can be used in an educational setting.

**Describe at least two (smaller, and preliminary) concrete results you have at this point.**

We realized the limitations of a dataset of reviews that are short and varied in style, grammatical precision, and content. While looking at the results of TF-IDF-inspired adjective-category mappings tells us that adjectives have some discerning power because they are in relative abundance, adjectives alone would be a little weak unless we had a larger dataset. Adjectives tended to be value-based (i.e. “great”) rather than descriptive. Second, we tried extracting just noun phrases (adjectives followed by nouns). We hypothesized that adjectives paired with nouns would give us more information to make judgments about the category of a review. In this subset of all nouns and adjectives, however, we found that a majority of the nouns that refer specifically to a category (i.e. “room”), rather than the hotel as a whole, is scarce (in the range of 1 to 6 occurrences). There are a few nouns in this set that appear in high abundance, but they describe hotels as a whole. We could use the said scarce nouns to discriminate between categories, but need to be wary of noise.