Ryan Berry

Nov 7th, 2021

CS410 – Text Information Systems

Technology Review: LingPipe and Cross-Domain Sentiment Analysis

Sentiment Analysis is defined as the use of natural language processing (NLP), machine learning, and other data analysis techniques to analyze and derive objective quantitative results from raw text. The most common applications of Sentiment Analysis include: social media monitoring, customer support/feedback, and product analysis. LingPipe, which is a tool kit for processing text using computational linguistics, offers a Sentiment Analysis framework capable of tackling those applications. I believe, however, that this technology could be used in conjunction with different domains. A unique factor of Sentiment Analysis is that its output is not limited to discrete data. It can also provide continuous values. I am particularly interested in exploring how these continuous values could bridge gaps between text, audio, and video.

To assist with my experimentation, I use the LingPipe tool kit. This platform-independent, Java software, includes 20 tutorials for different computational linguistic tasks. While I'm mostly interested in the Sentiment Analysis tutorial, it's worth noting that the library supports word sense disambiguation and Chinese word segmentation, which are notorious challenges in text analysis. This speaks to the depth and breadth of the software.

LingPipe's Sentiment Analysis tutorial focuses on movie reviews. It breaks up the analysis of the reviews into polarity and subjectivity. Step 1 removes sentences that are deemed to be objective (factual), while step 2 categorizes these sentences based on their polarity (positive/negative). But how does all this work? Diving deeper into step 2, we start out

with a classifier and a dataset including 1,000 positive and 1,000 negative movie reviews from IMDB's archives. The classifier is initialized as a "Dynamic Language Model Classifier" (DLMC), which operates as a multivariate estimator with Laplace smoothing. For training, the DLMC accepts a CharSequence sentence along with the a String representing the sentence's correct category as input. After training, the classifier can be queried with a new CharSequence input, providing a Classification object in return. The Classification object maintains the best result for categorizing an input. When training and classifying for polarity, the classification boils down to two categories: positive and negative. For the purposes of my experimentation, this value is not particularly useful. Ideally, the classification value would fall into a continuous range of 0 to 1, where 0 represents 100% negative and 1 is 100% positive. My understanding is that in order to generate these continuous values, I would need to avoid using a classifier and instead leverage some form of linear regression. Assuming that I am able to generate polarity values between 0 and 1, I can continue my experimentation with cross-domain sentiment analysis.

Now I pose the question, "What happens when we map this type of continuous value to the audio and video domains?" The continuous polarity spectrum that I described above is similar to the spectrum that represents color. One ranges from 0 to 1, while the other spans 0 to 255. While there is no accepted answer for, "What is the color of angry?" maybe we can generate one. After determining the color of angry, we could look at the opposite end on a color wheel and decide that this color represents "happy." Now envision an entire text document boiled down into RGB values. What is the use of this indexed form? Is it just art? Does it reveal any patterns? Could it be used as a practical visual aid? I am interested in finding out. Similarly, but within a different domain, I am interested in understanding how continuous

sentiment analysis representations can map to music. How can the text sentiment spectrum map to the musical note spectrum? My final project will develop a music generator to explore exactly this.

Works Cited

Tutorial - http://www.alias-i.com/lingpipe/demos/tutorial/sentiment/read-me.html

Intro quote - https://monkeylearn.com/sentiment-analysis/

SA applications - https://monkeylearn.com/blog/sentiment-analysis-applications/