Sentiment Analysis of Travel Reviews

Vanya Sehgal | Advisor: Professor Carol Romanowski | Rochester Institute of Technology

Problem Statement
Implement a sentiment analysis algorithm on travel reviews to classify the polarity of the reviews.

Importance
- As a traveler, one could quickly search for negative and positive reviews, based on that quick decision making can be done.
- A lot of research work is being done to determining how negative comments affects a person on social media.
- Helpful for businesses as well to quickly identify the issues reported.

Data Cleaning
- Stop word removal
- Conversion of abbreviated words into their actual meaning.
- Replacing emoticons with the emotion they represent.
- Removal of spam data.
- Replacing punctuations like “!!!!!!” with “!”.

Algorithm
- **Step 1: Tokenization**
  This is the first step towards achieving the goal. In this step the review text is split into tokens like punctuations, words and number.
- **Step 2: Speech Tagging**
  For every word in the review text a tag is produced that an-notates the role played by each word, whether it’s a noun, an adjective, an adverb or a verb.
- **Step 3: Word Sense Disambiguation**
  Word Sense Disambiguation (WSD) aims to deduce the meaning of each word in the context. Here, WSD implies determining the Synset for that word. Synsets are groups of synonyms with examples of its operation and describes correlation among these synonym sets.
- **Step 4: SentiWordNet Interpretation**
  Based on the Synset derived in the step above, the score of the word can be searched in the SentiWordNet.
- **Step 5: Sentiment Orientation**
  The negative and positive score for each word is calculated and summed up distinctly.

Data Collection
The data was collected with the help of APIs provided by Expedia.com. Hotel reviews of approximately 25 cities have been collected. The image below shows the JSON output of the data.

Only two attributes were used **ratingOverall** and **reviewText**.

Conclusion
- Number of positive remarks were 57%, Negative Remarks as 42% and Neutral remarks as 1%.
- It’s noted that reviewers were more interested in sharing their experiences if it was either negative or positive.
- Accuracy measured out to ~83%.

Future Work
- Work on a hybrid approach to optimize and improve the accuracy.
- The implemented algorithm can be used as a part of web browser extension, which could help future travelers to preview the travel reviews on one click.
- This extension could also help business owners in directly addressing to the negative reviews and work on improving it.

Experimental Results

**Fig 1. 5-step Algorithm**

**Fig 2. Sentiment Analysis Result**

**Fig 3. Word Cloud**
The Word cloud depicts the frequency of words found in data set. Bigger the text size, higher the frequency. Here, the larger texts are mostly positive, thus indicating more positive reviews.

References
- Gokulakrishnan, B; Priyanthan, P; Ragavan, T; Prasath, N; Perara, A. 2012. Opinion Mining and Sentiment Analysis on a Twitter Data Stream.