Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews
Peter D. Turney

Heman Baral

I. Description of the Study

The purpose of this study is evaluating semantic orientation. The average semantic orientation of the phrases contains adjectives or adverbs. A phrase has a positive semantic orientation when it has good associations and a negative semantic orientation when it has bad associations. The semantic orientation of a phrase is calculated as the mutual information between the given phrase and the word excellent minus the mutual information between the given phrase and the word poor. By looking at Epinions reviews of automobiles, banks, movies, and travel destinations The algorithm achieves an average accuracy for recommended.

II. Methods and Design

In a search engine, we search for some information in the query expected information and found the matches. It divides the query into fractions and matches recommend phrase. Search engines report the summary data with an unsupervised learning algorithm. The algorithm takes a written review as input and produces a classification as output. This algorithm first uses part-of-speech tagger to identify phrases in the input text that contain adjectives or adverbs then estimate the semantic orientation of each extracted phrase. A phrase has a positive and negative semantic orientation. Based on the average semantic orientation of the phrases extracted from the review by assigning a class. If the average is positive then review recommends the item otherwise the item is not recommended. The PMI-IR algorithm estimates the semantic orientation of a phrase and measures the similarity of pairs of words or phrases. A phrase is assigned a numerical rating by taking the mutual information between the given phrase. The word excellent and subtracting the mutual information between the given phrase and the word poor. The direction of the phrases semantic orientation positive or negative, based on the sign of the rating. Hatzivassiloglou and McKeown have also developed an algorithm for predicting semantic orientation. Designed for isolated adjectives, rather than phrases containing adjectives or adverbs.

III. Analysis

The PMI-IR algorithm extracts two consecutive words, where one member of the pair is an adjective or an adverb and the second provides context. Comparing its similarity to a positive reference word excellent and negative reference word poor calculate the semantic orientation of a given phrase.positive or negative numerical rating indicates the strength of the semantic orientation based on the magnitude of the number. Hatzivassiloglou and McKeown use a four-step supervised learning algorithm to infer the semantic orientation of adjectives from constraints on conjunctions. All conjunctions of adjectives are extracted from the given corpus and supervised learning algorithm combines multiple sources of evidence to label pairs of adjectives as having the same or different semantic orientation. A clustering algorithm processes the graph structure to produce two subsets of adjectives. Mainly different-orientation links and links inside a subset are mainly same-orientation links. Positive adjectives tend to be used more frequently than negative adjectives, the cluster with the higher average frequency is classified as having a positive semantic orientation.

IV. Results

The average Semantic Orientation recommends by the review. The classification algorithm is evaluated on 410 reviews from Epinions, randomly sampled from the domains of automobiles, banks, movies, and travel destinations. From these reviews, 170 are not recommended and the remaining 240 are recommended. The algorithm achieves an average accuracy of 74 percent while the baseline of accuracy is 59 percent. The clustering algorithm classifies adjectives with accuracies ranging from 78 percent to 92 percent, depending on the amount of training data that is available.

V. Limitations

The limitations of this work include the time required for queries and the level of accuracy that was achieved. Thumbs up or down system for generating sentiment timeline tracks online discussions about movies and displays a plot of the number of positive sentiment and negative sentiment messages over time but messages are classified by looking for specific phrases that indicate the sentiment of the author towards the movie. The movie aspect is not the same in every place and also they took a smaller sample size. An isolated adjective may be insufficient context to determine semantic orientation. Based on the presence or absence of specific words

could supplement average semantic orientation in a supervised classification system and it could yield higher accuracies.

VI. Significance

The review classification is based on an average which is might be quite resistant to noise in the semantic orientation estimate for individual phrases but a semantic orientation estimator can produce a better classification. It might benefit from more sophisticated statistical analysis to apply a statistical significance test to each estimated SO. There is a large statistical literature on the log-odds ratio, which might lead to improved results on this task. In this section discuss the significance of the work.

VII. Conclusion

A simple unsupervised learning algorithm takes a written review as input and produces a classification as output which makes our life easy. The searching engine divided into parts-of-speech and identify the target. For a small part of information, it followed semantic orientation and divided into good and bad associations. The PMI-IR algorithm is employed to estimate the semantic orientation of a phrase and given phrase is calculated by comparing its similarity to a positive reference word excellent and negative word poor. Hatzivassiloglou and McKeown have also developed an algorithm for predicting semantic orientation. Their algorithm performs well, but it is designed for isolated adjectives, rather than phrases containing adjectives or adverbs. The semantic orientation is calculated using the natural logarithm base e, rather than base 2.

References

[1] Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. [Cs/0212032] Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews, Peter D. Turney, 11 Dec. 2002, arxiv.org/abs/cs/0212032..