

# Text Sentiment

Sentiment analysis is the process of systematically identifying and quantifying the sentiment expressed in a text.

## Lexicon-based / dictionary-based approach

A common method is the lexicon-based approach, where each word is assigned a sentiment score, and the overall sentiment of a sentence is calculated as a weighted average of the words within it. Here, we adopt the approach used by `sentimentr` (Rinker, 2017), which utilizes a lexicon of polarized words (Hu & Liu, 2004; Jockers, 2017) and adjusts these scores with valence shifters.

The lexicon-based sentiment analysis begins with tokenization, where each paragraph ( $p_i$ ) is broken down into individual sentences ( $s_1, s_2, \dots, s_n$ ), and each sentence ( $s_j$ ) is further decomposed into a sequence of words ( $w_1, w_2, \dots, w_m$ ). Thus, each word can be represented as  $w_{i,j,k}$ . For instance,  $w_{2,3,1}$  refers to the first word in the third sentence of the second paragraph.

Next, the words  $w_{i,j,k}$  in each sentence are compared against a dictionary of polarized words. Positive words ( $w_{i,j,k}^+$ ) and negative words ( $w_{i,j,k}^-$ ) are assigned scores of +1 and -1, respectively. The context surrounding each polarized word is then analyzed, identifying neutral words ( $w_{i,j,k}^0$ ), negative modifiers ( $w_{i,j,k}^n$ ), amplifiers ( $w_{i,j,k}^a$ ), and de-amplifiers ( $w_{i,j,k}^d$ ). The sentiment score of each word is first weighted by its own score, and then further adjusted based on the function and quantity of valence shifters within its context. The sentiment score of the text is the average sentiment score of all words in the text.

## AI-based sentiment analysis

The Korn Ferry Institute's AITMI team made `sentiment.ai` for researchers and tinkerers who want a straight-forward way to use powerful, open source deep learning models to improve their sentiment analyses. Wiseman et al. (2022) packed the method in an R package `sentiment.ai` that can produce the sentiment of text and it outperforms many other methods.

The method is based on the Universal Sentence Embedding that embeds a text into a 512 by 1 vector. Then, it build a model between the embedded vector and the labels between the text for prediction.

# Online app

We have developed online apps for both dictionary-based and AI-based sentiment analysis. We created a video to show how to use the AI-based methods to get the sentiment of a text variables. The obtained sentiment score is saved as a new variable in the data set that can be used in further data analysis.

---

Revision #4

Created 4 July 2024 20:01:37 by Admin

Updated 24 October 2024 20:56:05 by Admin