

Sentiment Analysis

Yingfei Zeng

Advanced Topics in Machine Learning and Data Mining

February 14, 2018

Sentiment analysis (or opinion mining) plays an important role in the field of natural language processing and information retrieval. It's a text categorization task that extracts the positive or negative orientation of the writer from the text.

The traditional supervised learning process of sentiment analysis is to take a dataset and preprocess it at first, which often contains tokenization and stop words removal. Then opinion word extraction and labeling are conducted. POS tagging is always employed in this step to classify tokens into their respective genres, and extract the words that depict the emotion of the reviewer and label them as positive or negative based on the sentiment they represent. After that, classification algorithms can be used to achieve polarity determination. Some good algorithms include Naïve Bayes, Support Vector Machines (SVMs), Maximum Entropy, and so on. For instance, Ghaith Abdulsattar A. Jabbar Alkubaisi et al. propose an improved model with Hybrid Naïve Bayes Classifiers (HNBCs) that achieved accuracy equals 90.38%.

However, traditional supervised learning relies heavily on feature selection, which requires much domain-specific knowledge. Moreover, traditional methods regard a document as a bag of its words (BoW), which ignore word order and will lose some semantic and syntactic properties of the text. Due to these reasons, deep learning methods have been widely applied to sentiment analysis in recent years, which are able to learn multiple layers of features of the data automatically and take word order into consideration. For example, Rie Johnson et al. apply CNN to high-dimensional text data and achieve the accuracy of 92.86%.

Spotlight Question: What difficulties do you see in sentiment analysis?

Reading materials:

1. Alkubaisi, G. A. A. J., Kamaruddin, S. S., & Husni, H. (2018). Stock Market Classification Model Using Sentiment Analysis on Twitter Based on Hybrid Naive Bayes Classifiers. *Computer and Information Science*, 11(1), 52.
2. Zhang, L., Wang, S., & Liu, B. (2018). Deep Learning for Sentiment Analysis: A Survey. *arXiv preprint arXiv:1801.07883*.
3. Johnson, R., & Zhang, T. (2014). Effective use of word order for text categorization with convolutional neural networks. *arXiv preprint arXiv:1412.1058*.