

Sentiment Analysis of Chinese Microblog Comments: Comparison between Support Vector Machine and Long Short-Term Memory

Authors : Xu Jiaqiao

Abstract : Text sentiment analysis is an important branch of natural language processing. This technology is widely used in public opinion analysis and web surfing recommendations. At present, the mainstream sentiment analysis methods include three parts: sentiment analysis based on a sentiment dictionary, based on traditional machine learning, and based on deep learning. This paper mainly analyzes and compares the advantages and disadvantages of the SVM method of traditional machine learning and the Long Short-term Memory (LSTM) method of deep learning in the field of Chinese sentiment analysis, using Chinese comments on Sina Microblog as the data set. Firstly, this paper classifies and adds labels to the original comment dataset obtained by the web crawler, and then uses Jieba word segmentation to classify the original dataset and remove stop words. After that, this paper extracts text feature vectors and builds document word vectors to facilitate the training of the model. Finally, SVM and LSTM models are trained respectively. After accuracy calculation, it can be obtained that the accuracy of the LSTM model is 85.80%, while the accuracy of SVM is 91.07%. But at the same time, LSTM operation only needs 2.57 seconds, SVM model needs 6.06 seconds. Therefore, this paper concludes that: compared with the SVM model, the LSTM model is worse in accuracy but faster in processing speed.

Keywords : sentiment analysis, support vector machine, long short-term memory, Chinese microblog comments

Conference Title : ICACSIT 2022 : International Conference on Advanced Computer Science and Information Technology

Conference Location : Singapore, Singapore

Conference Dates : September 08-09, 2022