

A review on sentiment analysis model Chinese Weibo text

ABSTRACT

The technology of sentiment analysis about Chinese Weibo text is a complex and systematic model. In general situation, it includes 3 parts: data washing, word segmentation and feature extraction. Weibo text is an unstructured text and there are many non-standard contents in it. Therefore, it should be thoroughly data washing before feature extraction. Due to emoticon in Weibo text are very useful in sentiment analysis, thus, in data washing, all of Non-Chinese, with "@", "#" character should be removed except emoticon. In word segmentation, related algorithms can be divided into three categories: based on string matching, based on understand and based on statistics [1]. In feature extraction, the Lexicon-based Model, Machine learning Model and deep learning Model usually was used. Through literature search, the paper found that grammar characteristic in Chinese Weibo text was fully considered and solved by program of Lexicon-based Model, sentiment word, for example, adverb of degree, no word and all kinds of Chinese sentence patterns. But, due to characteristic of poor generalization, the performance of Lexicon-based Model in experiment is not good. Therefore, performance the model should be continued to improve. For traditional machine learning, there are 2 mainly aspects of innovation: Simultaneous classifier (Adoboost+SVM) and Improvement of classical classification algorithm. One worth noted is that the performance of the some improve classifier (SVM, P Naïve Bayes) has not been verified in Chinese Weibo classification. For deep learning, now, the innovation point is mainly focus on Convolution layer and input attention mechanism. For the next step, YuanHejin think should input ensemble learning and attention mechanism should be improve. LuXin argue that the recognition performance about irony sentence with context in Weibo needs to improve. GaoWeiju think that individual sentiment space for each user in EMCNN model should be build.