Weekly Summary

Jiulin Zong

Due:03/03/2021

1. Reading article *intro to gradable lexicon-based sentiment analysis*
2. Reading demo:
   1) Trying emoji on the demo: most utf-8 emoji can be properly analyzed
   2) Comparing with another one called Sentiment Polarity Analysis: using the same example provided by vader demo
      similar result but vader require less training time. Vader also has better performance in tricky examples in the demo
3. Reading the articles on advantage and limits of different sentiment analysis tools (related to the Sentiment Polarity Analysis)
   The accuracy rates are provided by articles.

   Dictionary based

   Accuracy rate: the accuracy rate is higher (more than 80%), with the increase of manual workload, the accuracy rate increases

   Advantages: easy to understand

   Disadvantages: heavy workload

   Based on K_ NN

   Accuracy: very low (60% - 70%)

   Advantages: simple thought, simple algorithm

   Disadvantages: low accuracy; memory consumption; time consumption

   Based on Bayes

   Accuracy: fairly good (70% - 80%)

   Advantages: simple, efficient, fast operation, good scalability

   Disadvantages: accuracy is not high, not practical

   Based on maximum entropy

   Accuracy: relatively high (above 83%)

   Advantages: high accuracy

   Disadvantages: long training time

   Based on SVM

   Accuracy: the highest (more than 85%)

   Advantages: high accuracy

   Disadvantages: training takes time