NTUSD-Fin: A Market Sentiment Dictionary for Financial Social Media Data Applications

Please cite the following paper if you use this dictionary in your research:
Chung-Chi Chen, Hen-Hsen Huang and Hsin-Hsi Chen. 2018. NTUSD-Fin: A Market Sentiment Dictionary for Financial Social Media Data Applications. In *Proceedings of the 1st Financial Narrative Processing Workshop*, 7 May 2018, Miyazaki, Japan.

 It contains 8,331 words, 112 hashtags and 115 emojis. The official download site is http://nlg.csie.ntu.edu.tw/nlpresource/NTUSD-Fin/.

There are 8 parts for each token (word, hashtag, or emoji), and they are saved in json format.
“token”: word, hashtag, or emoji
“bull_freq”: frequency in bullish set.
“bear_freq”: frequency in bearish set
“bull_cfidf”: collection frequency in bullish set.
“bear_cfidf”: collection frequency in bearish set.
“chi_squared”: chi squared test result of the token
“market_sentiment”: calculated by bullish PMI minus bearish PMI.
“word_vec”: 300-dimension word vector.
E.g.

{
'token': 'buy',
'bull_freq': 14489,
'bear_freq': 1592,
'bull_cfidf': 61.539806954702385,
'bear_cfidf': 52.32250663139482,
'chi_squared': 14711.705215251208,
'market_sentiment': 0.5961743093876137,
'word_vec': [0.0928284227848053, -0.10893399268388748, 0.12348346412181854, … , -0.01443735882639885]
}