Accepted Manuscript

Sentiment analysis of Chinese micro-blog text based on extended sentiment dictionary

Shunxiang Zhang, Zhongliang Wei, Yin Wang, Tao Liao

PII: S0167-739X(17)30783-5
DOI: https://doi.org/10.1016/j.future.2017.09.048
Reference: FUTURE 3704

To appear in: *Future Generation Computer Systems*

Received date : 27 April 2017
Revised date : 19 September 2017
Accepted date : 20 September 2017

Please cite this article as: S. Zhang, Z. Wei, Y. Wang, T. Liao, Sentiment analysis of Chinese micro-blog text based on extended sentiment dictionary, *Future Generation Computer Systems* (2017), https://doi.org/10.1016/j.future.2017.09.048

Sentiment Analysis of Chinese Micro-Blog Text based on Extended Sentiment Dictionary

Shunxiang Zhang , Zhongliang Wei, Yin Wang and Tao Liao

Abstract: Micro-blog texts contain complex and abundant sentiments which reflect user's standpoints or opinions on a given topic. However, the existing classification method of sentiments can not facilitate micro-blog topic monitoring. To solve this problem, this paper presents a sentiment analysis method for Chinese micro-blog text based on the sentiment dictionary to support network regulators' work better. First, the sentiment dictionary can be extended by extraction and construction of degree adverb dictionary, network word dictionary, negative word dictionary and other related dictionaries. Second, the sentiment value of a micro-blog text can be obtained through the calculation of the weight. Finally, micro-blog texts on a topic can be classified as positive, negative and neutral. Experimental results show the effectiveness of the proposed method.

Key Words: micro-blog text; sentiment dictionary; sentiment analysis; sentiment value

1 Introduction

With the rapid development of social network, a growing number of people express their views to some hot topics through the network. Therefore, micro-blog has become the one of the most common data platforms for users. On the micro-blog platform, the published texts by users are growing exponentially. These texts relate to society, life, science and technology, entertainment and other fields. In the era of big data, these texts which contain the user's subjective sentiment information have become pretty valuable information resources. An increasing number of organizations are making use of the sentiment information on the web for decision-making.

However, how to quickly extract the key and valuable sentiment information from massive micro-blog texts has become the focus of the current research. The expansion of the information network leads to relying on the manual way to collect and deal with the huge amount of information alone is too difficulty. Therefore, we require a new kind technology to generate valuable information quickly. The speed and accuracy of computer data processing are much more efficient than manual processing. So, the sentiment analysis technology that depends on computer technology is born. The sentiment analysis of micro-blog is highly concerned by a lot of scholars, mainly because it has many different characteristics with the ordinary text that with more complete textual content and longer length. Such as micro-blog texts are real time, short, diversity of information elements and so on. These features of micro-blog have also brought new challenges to the sentiment analysis. Through the analysis of the characteristics of micro-blog, several crucial questions about micro-blog sentiment analysis has been presented:

(1) Micro-blog text has flexible language, how to make use of these new network items and expressions in micro-blog text?