



4th Information Systems International Conference 2017, ISICO 2017, 6-8 November 2017, Bali, Indonesia

Effects of Word Class and Text Position in Sentiment-based News Classification

June Ling Ong Hui, Gan Keng Hoon*, Wan Mohd Nazmee Wan Zainon

School of Computer Sciences, Universiti Sains Malaysia, 11800 Penang, Malaysia

Abstract

In news domain, sentiments captured in the form of sentiment labels (emoticon) give a quick feedback of reactions towards the contents of the news. As these reactions are valuable indicators for social and political well beings, we are motivated to automate the classification of news texts based on these indicators, e.g. happy, sad, angry, amused etc. Unlike other review texts that contain more explicit words which can be interpreted directly for sentiment classification, news texts mostly report facts and figures. This resulted in needs to identify whether contents of news can be exploited for classification or otherwise. Hence, in this work, a study is conducted to analyze and determine the relevant key parts of news contents that can be used for sentiment-based classification. Two criteria, i.e. text Part of Speech and text position, which could possible influence the training of the classifier are studied. Evaluations are conducted on the collection of 250 English news texts labelled with sentiments from sentiment voting system. The results for sentiment-based category has recorded F score of 0.422 whereas for polarity-based category has recorded F score of 0.837. The study has shown that when finer categories (e.g. happy, sad etc.) are used, the inspected criteria are less effectively; however, when these categories are based on polarity orientations, the outcomes show potentials of the proposed criteria especially for text positioned at headlines and text using adjective words.

© 2018 The Authors. Published by Elsevier B.V.

Peer-review under responsibility of the scientific committee of the 4th Information Systems International Conference 2017.

Keywords: News Sentiment; Text Classification; Text Analysis

1. Motivation

Publication of news involves events, stories, reports from all walks of life normally triggers reactions from the readers. This trigger does not limit to paper-based news medium but also digital as well. People tend to have their

* Corresponding author. Tel.: +604-6534634.
E-mail address: khgan@usm.my

feeling expressed by talking and debating the topic of interest at the coffee shop on in the neighborhood during the days when the internet has not been widely used yet. Such scenario of offline discussion is often emotionally oriented and still can be seen anywhere until today. In this digital age, the online platform provides another alternative for the expression of thoughts and feelings. Some of these online platforms are social media sites and blogs such as Twitter, Facebook and Blogger where the online users usually express their feelings in text or comment.

Among these, one of the interesting ways for gathering feedback from online users is via the sentiment voting system. As a news article may provoke mix emotions and different readers may feel differently, the sentiment voting features provided in some websites are believed to have contributed greatly in understanding the readers' reaction after reading the news. This special feature captures the reader's reaction or attitude towards a situation by using emotion representation. Fig. 1 shows a news excerpt taken from TheStar online that shows the number of votes for each sentiment label. The six reactions captured are Happy, Inspired, Amused, Sad, and Annoyed, with three positive and three negative sentiments. Number of votes generated by users for a particular sentiment suggest that the news has higher tendency to provoke such emotion after reading it.

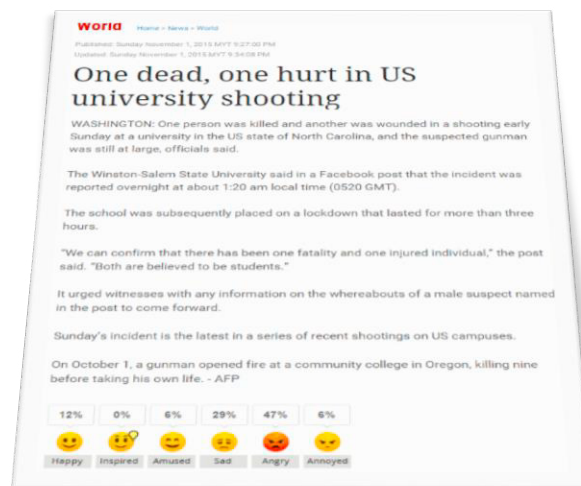


Fig. 1. News article with sentiment voting (The Star Online <http://www.thestar.com.my/>).

As emotions are often one of the influencing factors during decision making, knowing what others think and feel is essential for various group of people in the society. Marketer and advertiser can track public responses of an advertising campaign; politician and public relation officials can get insights from public emotions reacted on certain event especially on the political situation of a country; stock traders are able to capture the market behaviors for stock trading etc. Hence, this results in the needs to improve existing news categorization based on emotional labels. Handling emotion-based categories is challenging, and requires different treatment in terms of contents analysis compared to related works that have mostly focused on classification based on general categories like sports, politic, entertainment etc. Contents analysis based on measures like term frequency, n-grams, may not be able to differentiate those sentiments associated terms then general keywords. Thus, we believe that weighting should be given to certain type of terms so that they can better reflect the concept of emotion-based category. For example, news headlines could be more significant compared to news bodies in catching readers' attention, hence useful to be associated to the effect of an emotion. Likewise, type of word classes may also have different effect in terms of arousing the emotion of reader which resulting to a vote decision.

As such, in this paper, we focus on the analysis of textual contents in the setting of sentiment-based news classification. Our research scope is set to discover appropriate word classes (nouns, verbs, and adjective) as well as the type of news contents (headlines, first paragraph, last paragraph) in learning the model for classifying emotion label for news.

Download English Version:

<https://daneshyari.com/en/article/6901028>

Download Persian Version:

<https://daneshyari.com/article/6901028>

[Daneshyari.com](https://daneshyari.com)