Jun Chen
CSC6350
Software Engineering
2020Fall

Summary of clustering depressed and anti-depressed keywords based on a Twitter event of Srilanka bomb blasts using text mining methods

Problem Statement:

Nowadays, many people prefer to use Twitter or Facebook to share their feelings, emotions, location, and daily life. Twitter provides a maximum of 140 characters as a hashtag keyword. The authors what to figure out whether keywords are related to the event. Also, the authors want to find out the relationship between depressed keywords set and anti-depressed keywords set.

Proposed Solution:

The authors use Twitter API to collect tweets from Dambulla. Then, the authors also pick the top seven keywords from Kessler. The authors select the tweets based on the seven keywords and seven the antonyms for the seven keywords. The authors also compare two-period tweets. One is from April 14, 2019, to April 30, 2019. The other is from May 14, 2019, to May 25, 2019. The authors choose to use the Gradient-based method to analyze the datasets. The formula is $C_i = (ni/\Sigma ni)/(nb/\Sigma nb)$. In this formula, i represents each keyword. The authors also calculate Learning Quotient value[Qi] by using the formula $Q_i = P_i * ln(P_i), P_i = S_i / (\sum_{i=1}^{n} S_i + \sum_{t=1}^{n} DS_i)$ and calculate the keyword contribution factor of each set by using the formula $A = Ai * \sum_{i=1}^{n} Ai / (\sum_{i=1}^{n} Ai + \sum_{i=1}^{n} Bi)$. Then, the authors use text mining methods to calculate the Confusion Matrix by using the formula $AC = (TP + TN)/(TP + FN + FP + TN)$. The authors also calculate Association Factor by using the formula $AF = (A * B)/(A\text{^}2 + b\text{^}2 - A * B)$.

Results:

By comparing the results and see the figure, the authors find out that on the bombing day, the keywords combine near very low value. On all days, depressing tweets and anti-depressed tweets have a similar probability. However, the bombing tweets changes during the bombing day. The authors conclude that people post depressive tweets more than non-depressive tweets during the bombing day, and people who use the bomb as the hashtag post more depressive tweets.

Evaluation:

The advantage of this paper is that the authors use a lot of methods to get the results and prove that tweet keywords can be divided into two groups. It is easy for readers to understand the authors' opinion and ideas.

The disadvantage of this paper is that the authors just consider the keywords can be divided into two groups by using a lot of methods. Besides, the authors do not consider the difference between each method and do not show the advantage and disadvantages of the methods.

Synthesis:

The authors should consider whether people's tweets use other normal keywords to show depression. Besides, the bombing event may influence for a long time. The

period should consider longer. April and May are too close. The comparison may not be accurate. Besides, it is no doubt that if people suffer the bomb attack, they will feel depressive. The authors should consider whether other factors can cause people to post depressive tweets.

Reference:

S. T. Sadasivuni and Y. Zhang, "Clustering Depressed and Anti-Depressed keywords Based on a Twitter Event of Srilanka Bomb Blasts using text mining methods," 2020 IEEE International Conference on Humanized Computing and Communication with Artificial Intelligence (HCCAI), Irvine, CA, USA, 2020, pp. 51-54, doi: 10.1109/HCCAI49649.2020.00014.