BJUT at TREC 2020 Incident Streams Track

Hesong Wang*, Zhen Yang
College of Computer Science, Beijing University of Technology, China
1124759994@qq.com

Abstract

In this paper, we will continue to use the new method in the 2019 version to continue the work of the 2020 TREC Incident Streams System task[1]. Social media has become an indispensable part of human life, such as Twitter, Weibo and so on. When natural disasters occur, such as fires, earthquakes, flash floods, tsunamis, mudslides and other natural disasters or shootings, robberies and other emergencies, if only through media reports, the time of the event will be very slow, leading to some preventable loss. People like to post disaster situations or events on social media. The purpose of the task is to filter such natural disasters or emergencies by classifying the text on twitter. Similarly, each tweet is prioritized and the tagged information is reported to the relevant personnel according to different priorities. Let the staff know about the progress of the incident to help. This article will introduce the framework and methods of the classification system, as well as the experimental results.

Introduction

In the social network , people like to use social media to share and record their lives. Such as Wechat, Weibo, Twitter, Instagram, Facebook, etc. Twitter is one of the most popular social networking platforms, with millions of tweets a day. These twitter include not only daily life, but also natural disasters such as fire, earthquake, mountain flood, tsunami, debris flow, typhoon, or emergencies such as shooting and robbery. If we can carry out effective data mining and monitoring of these information, it will provide great help for the rescue work of relevant personnel. Based on this, TREC 2020 incident streams task is based on different events, such as: Request (request rescue, provide service, request useful information), CallToAction (volunteer, move people, donation), Report (potential threat, weather), Other (sentiment, discussion), and the importance of each tweet (Low, Medium, High, Critical). According to the corresponding categories, different score calculation rules are defined. The importance of information is determined by weighting the score corresponding to information classification and the score output corresponding to importance classification. The second section introduces the classification method of this topic, the third section gives the experimental results, and the fourth section summarizes.

Incident Streams System Framework

This chapter will focus on the design of classification system used to complete tasks.The framework of the model is shown in Figure 1.The system consists of data preprocessing module, training module and prediction module.

- Data preprocessing module
  1.The original tweet contains a lot of noise information, such as web links, non-english characters such as @, RT, etc, which will have a negative impact on the subsequent model training. The methods used include regular expression, stop words, part of speech restoration, sentence Association, OOV dictionary to standardize the abbreviations in the tweet, and data standardization to preprocess.
  2. We use the bert pre-trained model for text classification.
- Prediction module

$$f(x) = \frac{\sum_{i=1}^{|O(x)|} O^i(x) + G(x)}{A + |O(x)| \times B} \tag{1}$$

The parameter x represents a tweet, and $f(x)$ represents each tweet combined with information classification and alarm classification weighting, and finally outputs the importance score of the tweet.$f(x)$ ranges from 0-1(if $f(x)$ is higher than 1, the value is assigned 1).The higher the score, the more the tweet important. $O(x)$ represents the category of the tweet information classification, because the information classification is also multi-label multi-class, so $O(x)$ is multi-valued, $|O(x)|$represents the number of categories of tweet information classification. $o^i(x)$represents the value corresponding to the i-th classification in the tweet information classification from Table 1, the parameter A represents the mean value of all information classification values from Table 1, the parameter B represents the mean value of the alarm classification value from Table 2, and $G(x)$ represents the tweet alarm classification corresponding to value from Table 2, because the alarm classification is a single classification, so there is only one value.

Submitted Runs and Experiment Results

From the experimental results, we can see that our model has a good performance in predicting the importance, but the

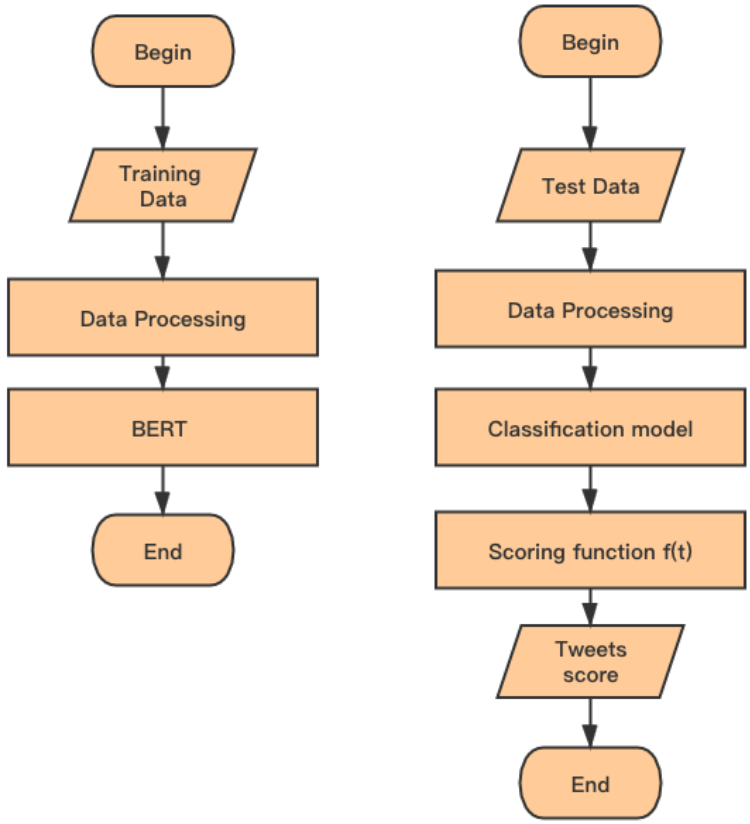

Figure 1: System Framework.

performance in information classification is not good. There are too many categories of information classification, short text has no context information, text expression is relatively flexible, can not express the mood of microblog author, corpus coverage is not comprehensive, these are the reasons that affect information classification. We will continue to participate next year. The next step is to improve the performance of information classification and strive for a good result.

Conclusion

This year we have used a new method, the performance has been greatly improved compared with last year, but the performance is not good in the classification of information. Short text multi-label and multi-category is very challenging. We will participate next year, hoping to Each indicator surpasses this year's performance.

References

1.Ning, Lu.;Hesong, Wang and Yang, Zhen. 2016. BJUT at TREC 2018: Incident Streams Track. In TREC.
2.Tianqi Chen ,Carlos Guestrin . XGBoost: A Scalable Tree Boosting System.
3.Badjatiya P , Gupta S , Gupta M , et al. Deep Learning for Hate Speech Detection in Tweets[J]. 2017.
4.Sundermeyer M, Schlüter R, Ney H. LSTM Neural Networks for Language Modeling[C]// Interspeech. 2012:601-608.
5.Wang, K., and Yang, Z. 2016. BJUT at TREC 2016: Real-time summarization track. In TREC.
6. Tan, H.; Luo, D.; and Li, W. 2016. POLYU at TREC 2016 realtime summarization. In TREC.
7. Lin, J.; Roegiest, A.; Tan, L.; McCreadie, R.; Voorhees, E.;and Diaz, F. 2016. Overview of the trec 2016 real-time summarization track. InProceedings of the 25th Text REtrieval Conference, TREC, volume 1.

Table 1: Information Quantitative Score

label	value
Request-GoodsServices	10
Request-SearchAndRescue	10
Request-InformationWanted	10
CallToAction-Volunteer	8
CallToAction-Donations	8
CallToAction-MovePeople	9
Report-FirstPartyObservation	4
Report-ThirdPartyObservation	4
Report-Weather	3.5
Report-Location	4
Report-EmergingThreats	8
Report-NewSubEvent	7
Report-MultimediaShare	2
Report-ServiceAvailable	5
Report-Factoid	3
Report-Official	5
Report-News	3.5
Report-CleanUp	3
Report-Hashtags	2
Report-OriginalEvent	3
Other-ContextualInformation	3
Other-Advice	4
Other-Sentiment	1
Other-Discussion	1
Other-Irrelevant	1

Table 2: Alarm Quantitative Score

level	value
Critical	10
High	7
Medium	4
Low	1

Run	nDCG@100	Info-Type F1 [Actionable]	Info-Type F1 [All]	Info-Type Accurary	Priority F1 [Actionable]	Priority F1 [All]
BJUT-run	0.4356	0.0266	0.0581	0.8321	0.1895	0.0959