

# News Text Classification Based on BERT and Relevant Promotions

Nan Shen, Qiqi Chen

Tsinghua-Berkeley Shenzhen Institute

## Abstract

Using different models and improved algorithms, we try to classify news texts. The results of models are compared and analyzed.

- Distributions of training and test data sets are basically the same, and both have obvious long tail phenomena.
- BERT, a pre-training and fine tuning model, can effectively improve accuracy of predictions.
- Parameter adjustment and model superposition are effective means.

## Introduction

Text classification technology is considered as one of powerful means to deal with mass information. It can effectively solve problems of information chaos and help users to accurately locate information. This work aims to:

- News text classification with different model algorithms and their promotion.

Attempt 1: classical basic model algorithms such as Word2vec, TextCNN, BERT.

Attempt 2: improvements and model integration based on BERT and LSTM.

- Comparison and analysis on model results.

## Data

THUCNews dataset.

- 740 million news documents. 20 thousand for training.
- 14 candidate categories: finance, lottery, real estate, stocks, home, education, technology, society, fashion, politics, sports, constellation, games and entertainment.

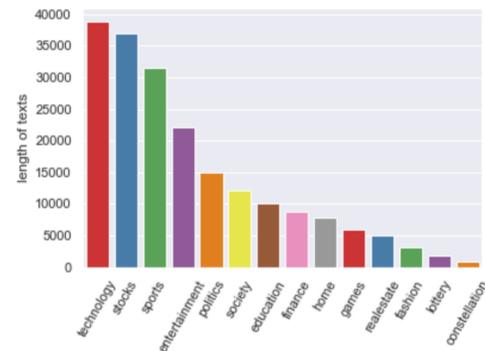


Figure 1: News category

- The distribution of categories of news is unbalanced. News of technology accounted for about 20 percent of the data.
- Distributions of length of documents and frequency of words are extremely right-skewed, suggesting a typical long-tailed distribution problem.

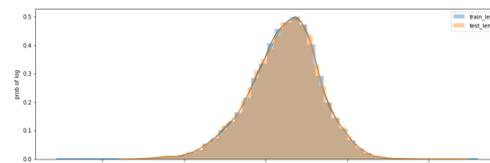


Figure 2: Distribution of logarithm of length of texts

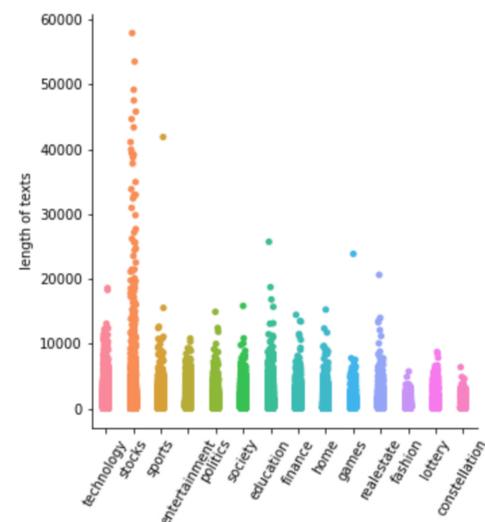


Figure 3: Length of texts of different categories

## Methods

We use Stratified K-fold to deal with unbalanced data. This method ensures similar distribution as original after sampling. This lowers the variance and enhances model capability.

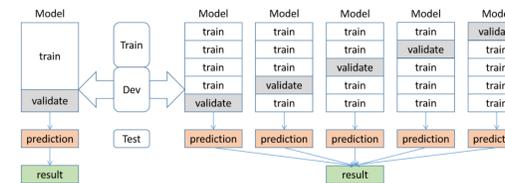


Figure 4: Stratified K-fold

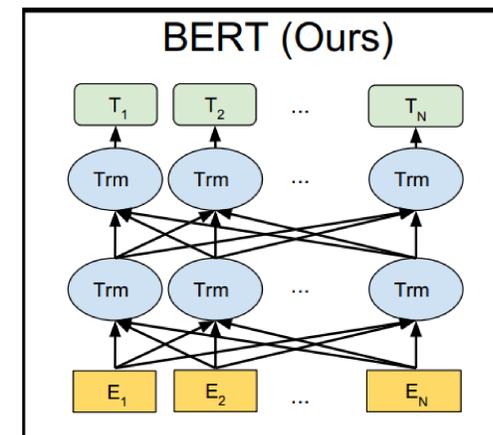


Figure 5: BERT

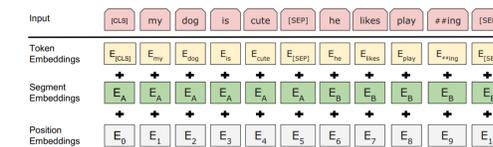


Figure 6: BERT Embedding

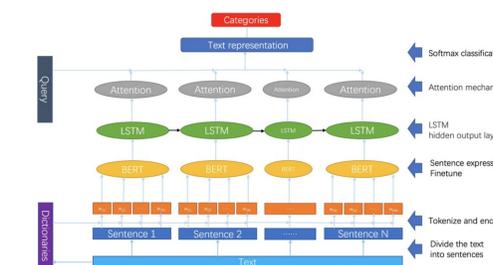


Figure 7: Modified BERT

## Results

Method	Accuracy
Count Vectors + RidgeClassifier	0.7415
TF-IDF + RidgeClassifier	0.8722
Fasttext	0.8752
LSTM	0.9499
TextCNN	0.9438
BERT	0.9534
BERT + LSTM	0.9659

## Conclusions

- Distributions of training set and test set are basically consistent, therefore results on validation set and test set are consistent.
- Multi-head attention mechanism and bidirectional encoding make unsupervised training of BERT more effective and allow BERT to construct a wider depth model.
- Parameter tuning and model integration are effective means of promotion.

## References

- Devlin, J. , Chang, M. W. , Lee, K. , Toutanova, K. (2018). Bert: pre-training of deep bidirectional transformers for language understanding.
- Cui, Yiming, et al. (2020). Revisiting Pre-Trained Models for Chinese Natural Language Processing. arXiv preprint arXiv:2004.13922
- <http://thuctc.thunlp.org/>

## Contact Information

- sn20@mails.tsinghua.edu.cn
- chenqq20@mails.tsinghua.edu.cn