

Shanghai-HK Interdisciplinary Shared Tasks

3rd Place Report Sharing for Task 1

Peng Xia 夏鹏

Soochow University 苏州大学

2022 年 6 月 6 日



養天正氣
法古今完人
楊永清題

- ① Introduction
- ② Methodology
- ③ Experiment
- ④ References

養天正氣
法古今完人
楊永清題

① Introduction

② Methodology

③ Experiment

④ References

養天正氣
法古今完人
楊永清題

Background

Nowadays, rumors tend to spread quickly and widely on the Internet, and automatically verifying rumors has become an urgent need for individuals and society. The task of rumor verification aims at classifying rumor cascade as true, false or unverified.

養天正氣
法古今完人
楊永清題

Task definition

Message roles are summarized into 4 categories, i.e. amplify, deny, clarify and null. Amplify indicates tweets that initiate new concerns or enlarge the discussion scale related to the social event. Deny means presenting doubt or rejection towards previous messages. Clarify introduces factual or substantial information. Other messages are left as null which means they are insignificant for rumor propagation or verification.

法
養
天
地
正
氣
古
今
完
人
楊
永
清
題

Dataset and Evaluation metric

- The extended dataset contains 1,929 cascades and 26,871 messages annotated. Train, validation, test (phase 1) and test (phase 2) sets are split randomly with a proportion of 7:1:1:1.
- We adopt macro F1 score is adopted as evaluation metrics.

表 1: Dataset Examples

cid	mid	pid	time	date	content
5249...	5249...	None	1413...	2014-10-22	@affa.: Kevin...
5249...	5249...	5249...	1415...	2014-10-23	@Linda: good for...

- ① Introduction
- ② Methodology
- ③ Experiment
- ④ References

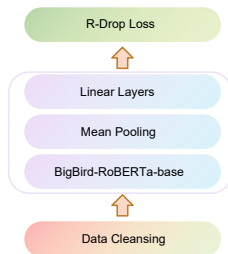
養天正氣
法古今完人
楊永清題

Data cleansing

- Convert the abbreviations in the text to the full version
- url, html tags and @ are removed
- special characters and numbers between null characters are removed as well

養天正氣
法古今完人
楊永清題

Modeling



- task is regarded as a text classification task
- use a sparse-attention Transformer encoder called BigBird as the backbone to obtain message representation.
- apply mean pooling to representations from encoder to acquire cascade representation

BigBird

BigBird, is a sparse-attention based transformer which extends Transformer based models, such as BERT to much longer sequences. Moreover, BigBird comes along with a theoretical understanding of the capabilities of a complete transformer that the sparse model can handle. It is a pretrained model on English language using a masked language modeling (MLM) objective. We use it through Transformers¹.

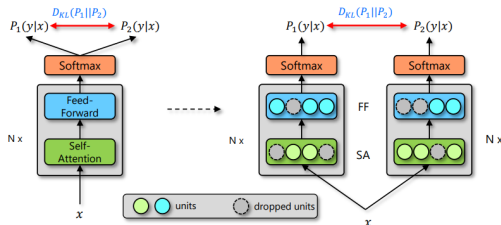
¹<https://huggingface.co/google/bigbird-roberta-base>

Improvements

- add extra linear layers
- apply R-Drop loss function which is a simple and efficient regularization technique based on Dropout
- replace KL-Divergence loss with MSE loss
- explore the influence of different forward times on model performance

法
養
天
地
正
氣
古
今
完
人
楊
永
清
題

R-Drop loss



- In the same step, for the same sample, we propagate forward twice.
- Due of the Dropout, we get two different probability distributions.
- By adding the KL divergence loss of the two distributions into the original cross entropy loss, the two distributions jointly carried out back propagation and parameter update.

$$L_R^i = \text{MSE}(y_1^i, y_2^i) \quad (1)$$

$$L_B^i = \frac{1}{2} \cdot (\text{MSE}(y_1^i, \hat{y}^i) + \text{MSE}(y_2^i, \hat{y}^i)) \quad (2)$$

$$L^i = \alpha \cdot L_R^i + (1 - \alpha) \cdot L_B^i \quad (3)$$

- Where y_1^i and y_2^i are the predicted values of the two forward of the sample, \hat{y}^i is the true value of the sample, and α controls the relative intensity of the two losses.
- By introducing the same sample and undergoing different Dropout of the same model, R-Drop is an optimization target whose probability of output should be as equal as possible, which is equivalent to the regularization constraint that makes all parameters of the model as equal as possible.

- ① Introduction
- ② Methodology
- ③ Experiment
- ④ References

養天正氣
法古今完人
楊永清題

- All experiments were tested on test1, but not test2.
-

System	macro F1 score
Baseline	1.194
RoBERTa	1.215
BigBird	1.263
+Extra Linear Layers	1.268
+R-Drop	1.313

表 2: Performance of different experiments

Anticipation and acknowledgement

- Other attempts: improve the pooling layer...
- explore the data itself: data augmentation...
- Thanks to the organization committee and all the staff!

養天正氣
法古今完人
楊永清題

- ① Introduction
- ② Methodology
- ③ Experiment
- ④ References

養天正氣
法古今完人
楊永清題

Zaheer M, Guruganesh G, Dubey K A, et al. 2020. Big bird: Transformers for longer sequences. *Advances in Neural Information Processing Systems*, 33: 17283-17297.

Opitz J, Burst S. 2019. Macro f1 and macro f1. *arXiv preprint*, arXiv:1911.03347.

Vaswani A, Shazeer N, Parmar N, et al. 2017. Attention is all you need. *Advances in Neural Information Processing Systems*, 30.

Wu L, Li J, Wang Y, et al. 2021. R-drop: regularized dropout for neural networks. *Advances in Neural Information Processing Systems*, 34.

Srivastava N, Hinton G, Krizhevsky A, et al. 2014. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1): 1929-1958.

Devlin J, Chang M W, Lee K, et al. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint*, arXiv:1810.04805.

法
養
古
天
今
地
完
正
人
氣
楊永清題

Thanks for your listening!

養天正氣
法古今完人
楊永清題