

CCF BDCI 文本相似度计算比赛

运行说明

1. 下载bert中文预训练模型: `wget`

```
https://storage.googleapis.com/bert_models/2018_11_03/chinese_L-12_H-768_A-12.zip
```

解压中放入 `./ckpt/` 中

2. 配置环境

- scikit-learn
- tqdm
- pandas
- numpy
- scipy
- jieba
- Keras
- tensorflow=1.14.1

代码结构

1.数据读取、预处理、特征提取

```
python code/extract_features.py
```

2.Bert模型训练

```
python code/bert.py
```

3.模型融合脚本

使用试例:

```
python code/vote_to_submit.py sub/bert*.csv sub/vote_bert.csv
```