# 周报，2021年03月15日

屈原斌
首都师范大学
ybqu@cnu.edu.cn

## 1 上周计划

1. [***] 生成实验更新
2. [***] 英文离题实验更新
3. [***] 中文测试集处理

## 2 上周计划执行情况

1. [✓]
2. [✓]
2. [×]

## 3 本周部分重点工作详述

### 3.1 英文数据集

#### 3.1.1 BERT生成模型更新

- 数据集：Xsum数据集，数据分布见表1
- 指标更新，见表2
- 生成样例，见图1
- 结论:
  - Bert生成模型指标较低，生成结果包含重复词和符号

| | 训练集 | 验证集 | 测试集 |
|---|---|---|---|
| **Xsum** | 214352 | 11333 | 11333 |

Table 1: 生成模型数据分布

| | | rouge_1 | rouge_2 | rouge_l |
|---|---|---|---|---|
| **BiLstm** | **MLE** | 0.4346 | 0.2290 | 0.4022 |
| | **+RL** | 0.4420 | 0.2364 | 0.4089 |
| **Bert-transformer** | | 0.1190 | 0.0134 | 0.1458 |

Table 2: 英文生成模型指标

正文：north korea has conducted a new intercontinental ballistic missile test , south korea and the pentagon say .  the missile reached an altitude of about 3 , 000 ##km ( 1 , 86 ##5 miles ) and landed in the sea off japan , the japanese national broadcaster nh ##k said .  it comes three weeks after north korea ' s first ic ##bm test .  in response , the us and south korean military conducted a live - firing exercise using surface - to - surface missiles , a us defence official said .  the missiles were fired into the ` ` territorial waters of south korea along the east coast , ' ' a us military statement said .  the latest north korean missile flew higher , further and for longer than the one in early july .  its launch has been condemned by a number of countries .  the test - the 14th carried out by north korea in 2017 - is the latest to be conducted in defiance of a un ban .  us president donald trump called it ` ` only the latest reckless and dangerous action by the north korean regime ' ' . jeffrey lewis , a nuclear non ##pro ##life ##ration expert at the middle ##bury institute of international studies in california , said that initial indications showed the latest missile had a range of about 10 , 000 ##km - far enough to strike the west coast of the united states and beyond .  the washington post reported that denver and possibly chicago could now be in north korea ' s range .  the latest missile was launched at 23 : 41 ( 15 : 41 gm ##t ) from an arms plant in ja ##gang province in the north of the country , the pentagon said . it is unusual for north korea to launch a missile at night - the significance is as yet unclear .  no missiles had been fired from ja ##gang province before , indicating a previously - unknown launch site is operational .  japanese chief cabinet secretary yo ##shi ##hide su ##ga said the missile flew for about 45 minutes - some six minutes longer than the ic ##bm tested in early july .  nh ##k said it reached an altitude of about 3 , 000 ##km - about 200 ##km higher than the previous ic ##bm .  it landed about 1 , 000 ##km from the launch site , the pentagon said .  ic ##bm ##s can reach altitudes well outside the earth ' s atmosphere .  using sharp tr ##aj ##ect ##ories with high altitudes allows north korea to avoid firing over neighbouring countries .  in the latest reaction : despite ongoing tests ,
预测：north missile missile missile korea missile missile : us missile missile
题目：north korea conducts new intercontinental missile test

Figure 1: Bert模型生成样例

### 3.1.2 离题实验更新(Accuray调阈值)

- 数据集划分：五折交叉验证
- 指标更新:
  - 方案一：基于题目排序方法（共70个参考题目），见表3、4
  - 方案二：基于相似度方法，见表5、6
  - 各个分数档下作文Recall，见表7（使用bert_cls）
- 实验结论：
  - Bertabs生成模型优于lstm
  - 使用Bert-whitening方法后指标下降，方案一上取[CLS]指标最优，方案二上最后一层表示取平均指标最优

  ？各分数档recall与分数基本成正比，即分数越高越容易识别(待分析)

| | | Accuracy | 离题 | | | 不离题 | | |
|---|---|---|---|---|---|---|---|---|
| | | | precision | recall | f1-score | precision | recall | f1-score |
| **bilstmabs** | 开发集 | 0.5867 | 0.6815 | 0.3194 | 0.3653 | 0.5863 | 0.7977 | 0.6618 |
| | 测试集 | 0.5413 | 0.5663 | 0.2817 | 0.3223 | 0.5544 | 0.7736 | 0.6301 |
| **bilstmabs_hc** | 开发集 | 0.5675 | 0.5874 | 0.4183 | 0.4055 | 0.5911 | 0.6642 | 0.5696 |
| | 测试集 | 0.5027 | 0.4461 | 0.3789 | 0.3427 | 0.5487 | 0.6203 | 0.5144 |
| **bertabs_cls** | 开发集 | 0.5904 | 0.5476 | 0.6053 | 0.5750 | 0.6341 | 0.5778 | 0.6047 |
| | 测试集 | 0.5015 | 0.4689 | 0.4855 | 0.4771 | 0.5322 | 0.5156 | 0.5237 |

Table 3: 方案一-生成模型离题指标

| | | Accuracy | 离题 | | | 不离题 | | |
|---|---|---|---|---|---|---|---|---|
| | | | precision | recall | f1-score | precision | recall | f1-score |
| **bert_cls** | 开发集 | 0.5542 | 0.8030 | 0.1592 | 0.1885 | 0.5407 | 0.8742 | 0.6582 |
| | 测试集 | 0.5361 | 0.7138 | 0.1690 | 0.1833 | 0.5428 | 0.8634 | 0.6469 |
| **bert_last1avg** | 开发集 | 0.5554 | 0.8024 | 0.1400 | 0.1723 | 0.5455 | 0.8970 | 0.6723 |
| | 测试集 | 0.5331 | 0.7021 | 0.1446 | 0.1648 | 0.5387 | 0.8782 | 0.6545 |
| **bert_last2avg** | 开发集 | 0.5566 | 0.8042 | 0.1352 | 0.1701 | 0.5466 | 0.9046 | 0.6764 |
| | 测试集 | 0.5316 | 0.6894 | 0.1347 | 0.1565 | 0.5372 | 0.8837 | 0.6568 |
| **bert_whitening(CLS)** | 开发集 | 0.5361 | 1.0000 | 0.0494 | 0.0941 | 0.5247 | 1.0000 | 0.6883 |
| | 测试集 | 0.5467 | 0.5676 | 0.0686 | 0.1224 | 0.5455 | 0.9553 | 0.6944 |

Table 4: 方案一-Bert取不同表示下离题指标

| | | Accuracy | 离题 | | | 不离题 | | |
|---|---|---|---|---|---|---|---|---|
| | | | precision | recall | f1-score | precision | recall | f1-score |
| **bilstmabs** | 开发集 | 0.5590 | 0.3471 | 0.2926 | 0.2766 | 0.5510 | 0.7485 | 0.5740 |
| | 测试集 | 0.5181 | 0.2959 | 0.2716 | 0.2430 | 0.5390 | 0.7440 | 0.5644 |
| **bilstmabs_hc** | 开发集 | 0.5663 | 0.6941 | 0.2978 | 0.2918 | 0.4540 | 0.7566 | 0.5673 |
| | 测试集 | 0.5099 | 0.4462 | 0.2770 | 0.2416 | 0.4245 | 0.7245 | 0.5342 |
| **bertabs_cls** | 开发集 | 0.5602 | 0.6250 | 0.2469 | 0.3540 | 0.5448 | 0.8588 | 0.6667 |
| | 测试集 | 0.5226 | 0.4610 | 0.2124 | 0.2908 | 0.5392 | 0.7877 | 0.6402 |

Table 5: 方案二-生成模型离题指标

### 3.2 中文数据集

- 处理数据（乱写检查）

## 4 下周计划

1. [***] 进一步分析实验结果，查看预测结果
2. [***] 完成中文数据处理

| | | Accuracy | 离题 | | | 不离题 | | |
|---|---|---|---|---|---|---|---|---|
| | | | **precision** | **recall** | **f1-score** | **precision** | **recall** | **f1-score** |
| **bert_cls** | 开发集 | 0.5940 | 0.7349 | 0.2885 | 0.3546 | 0.5805 | 0.8420 | 0.6800 |
| | 测试集 | 0.5358 | 0.5926 | 0.2641 | 0.2977 | 0.5508 | 0.7786 | 0.6283 |
| **bert_last1avg** | 开发集 | 0.5855 | 0.4970 | 0.3595 | 0.3861 | 0.5921 | 0.7737 | 0.6551 |
| | 测试集 | 0.5452 | 0.4279 | 0.3398 | 0.3452 | 0.5668 | 0.7275 | 0.6149 |
| **bert_last2avg** | 开发集 | 0.5880 | 0.5208 | 0.2879 | 0.3485 | 0.5860 | 0.8521 | 0.6850 |
| | 测试集 | 0.5494 | 0.4433 | 0.2418 | 0.2920 | 0.5550 | 0.8171 | 0.6519 |
| **bert_whitening(CLS)** | 开发集 | 0.5904 | 0.6818 | 0.1974 | 0.3061 | 0.5764 | 0.9222 | 0.7094 |
| | 测试集 | 0.5572 | 0.6024 | 0.1608 | 0.2538 | 0.5508 | 0.9065 | 0.6852 |

Table 6: 方案二-Bert取不同表示下离题指标

| **score** | **1.0** | **1.5** | **2.0** | **2.5** | **3.0** | **3.5** | **4.0** |
|---|---|---|---|---|---|---|---|
| **1**折 | 0.0000 | 0.0000 | 0.0000 | 0.0278 | 0.0465 | 0.0323 | 0.9713 |
| **2**折 | 0.0000 | 0.0000 | 0.5000 | 0.6667 | 0.5584 | 0.6667 | 0.3654 |
| **3**折 | 0.0000 | 0.0000 | 0.0000 | 0.0303 | 0.0227 | 0.0339 | 0.9749 |
| **4**折 | 0.0000 | 0.0000 | 0.0000 | 0.0303 | 0.0361 | 0.0370 | 0.9688 |
| **5**折 | 0.0000 | 0.0000 | 0.0000 | 0.0263 | 0.0465 | 0.0432 | 0.9770 |
| **avg** | 0.0000 | 0.0000 | 0.1000 | 0.1563 | 0.1421 | 0.1626 | 0.8515 |

Table 7: 各个分数档下作文Recall（bert_cls）