周报，2021年03月01日

屈原斌

首都师范大学

ybqu@cnu.edu.cn

1 上周计划

1. [***] 实验更新

2 上周计划执行情况

1. [✓×]

3 本周部分重点工作详述

3.1 英文数据集

3.1.1 分类模型更新

- 数据集：ICLE数据集，共70个主题（去除作文数小于20篇的主题），数据分布见表1

- 指标更新，见表2

- 实验结论：

    - 加入测试主题后指标下降较多（数据不平衡问题）

	训练集	验证集	测试集	合计
作文数	2838	322	388	3548

Table 1: 分类模型数据分布

	Accuracy	Percision	Recall	F1-score
HABiLstm	0.7139	0.6635	0.6663	0.6382
HABiLstm(加入测试主题)	0.6974	0.6064	0.5949	0.5768
BERT	0.9198	0.9079	0.8954	0.8914
BERT(加入测试主题)	0.8564	0.7653	0.7793	0.7602

Table 2: 英文数据分类指标

3.1.2 生成模型更新

- 使用Gigaword数据进行预训练后，在ICLE数据集上做进一步训练

- 数据集：Gigaword、ICLE数据集，数据分布见表3

- 指标更新，见表4

- 实验结论：

    - ICLE数据集上指标下降（ICLE正文和摘要/题目平均长度大于Gigaword）

	训练集	验证集	测试集
Gigaword	3803957	135129	135100
ICLE	8576	477	476

Table 3: 生成模型数据分布

	Rouge_l
seq2seq(Gigaword)	0.4263
seq2seq(+ICLE)	0.3811

Table 4: 生成模型指标更新

3.1.3 离题实验更新

- 数据集划分：五折交叉验证，数据划分见表5

- 指标更新，见表6

- 实验结论：

  - 添加测试集主题后指标下降（分类数据中可能存在离题作文）

Fold1	离题:切题 = 72:94
Fold2	离题:切题 = 87:79
Fold3	离题:切题 = 81:85
Fold4	离题:切题 = 76:90
Fold5	离题:切题 = 71:95

Table 5: 五折交叉验证数据划分

3.1.4 Attention可视化

- HABiLstm（未完成）

- Seq2seq生成模型

  - 见图1、2
  - 观察结果：
    * 部分生成长度较长
    * 关键词提取不准，attention值高的词都是一些连接词

4 下周计划

1. [***] 更新实验方案（聚类/回归方法）

2. [***] 整理更新中文实验结果

		离题			不离题		
		precision	recall	f1-score	precision	recall	f1-score
未添加测试集主题	habilstm	0.47324	0.556786	0.511149	0.530827	0.447184	0.484883
	bert	0.466495	0.627956	0.535201	0.533974	0.372536	0.438731
添加测试集主题	habilstm	0.546867	0.066731	0.118818	0.523113	0.947805	0.673808
	bert	0.50274	0.061962	0.110262	0.535799	0.945872	0.68404
doc2vec		0.462607	0.077407	0.132581	0.533329	0.921086	0.675471
seq2seq		0.469614	0.818419	0.595204	0.536597	0.190272	0.271571

Table 6: 英文数据集离题指标更新

正文: The criminals who break the law should experience that what they did was wrong and they should learn that it's stupid to do it again. But what is the best way to make the criminals realize what they've done and open their eyes. Punish or rehabilitate? Maybe combine the two? There is also a difference between what laws they break and how long the conviction is. The maingoal here is to make sure the criminals do n't do it again and how to achieve that . There is a difference between a person who rape and a person who do forgery. I believe the person who rape should be punished bad, but he also needs rehabilitation, but on the other hand the person who do forgery needs more rehabilitation than punishment. No matter what a criminal has done, he or she needs to be told one way or the other that this doesn't pay and the criminals need to realize it themselves. How to do that is a bigger problem. How to rehabilitate? The criminals are too comfortable in the prison here in Norway. They need to feel that they are in a prison and they've done something wrong and have to pay, but they also need rehabilitation. They have to find out a way to do rehabilitation that works and have a positive affect on people. They don't need to have a TV with all the channels. They don't have to eat very good food every day, all day. They don't have to have access to a computer with the Internet. They do n't have to live like the people on the outside. There is a reason that they're on the inside. They've done something wrong and they have to pay for that. I believe it's wrong that they all that luxury inside. It's a prison and it should not be a holiday place for the people less fortunate. It should be a place that people fear. It has to be a reason that the criminals do the crime over and over again even if they're caught. Why doesn't criminals fear the idea of going to prison for several years. It has to do with something like they think it's smart. Th at they can really make money out of it and some of them maybe even think that: "It doesn't matter if I get caught because it isn't so bad i nside. It's almost better than home except I can't go out whenever I want to." I think some criminals actually believe that crime pays. Seen it on TV and heard about criminals who did fine . So what we have to work on is how to make the criminals realize that what they're doing is wrong. Off course a civilized society should rehabilitate them, but can that be done without punishment of some kind? Also try to make th e prisons here in Norway more feared. Make the time that they have to spend inside, longer, at least for those who do something really bad, like rape! Rehabilitation is not to do the time by watching TV and eating good food. Crime doesn't pay. Make them suffer at least a little. Make them feel that what they've done is terribly wrong .

预测: The prison system is outdated; no civilized society should punish its criminals. It should rehabilitate

真实: The Prison System

Figure 1: 生成结果

be: 0.08094325661659241
be: 0.07732872664928436
learn: 0.0740908682346344
and: 0.040832605212926865
to: 0.036366891115903854
to: 0.026373134925961494
to: 0.024374978616833687
and: 0.024234553799033165
not: 0.022361954674124718
and: 0.020736213773489
to: 0.02053409069776535
and: 0.015920370817184448
and: 0.015883740037679672
and: 0.015434002503752708
and: 0.0151992086321155
be: 0.015114334411919117
and: 0.01299965102225542
and: 0.012923887930810452
they: 0.010417995974421501
also: 0.010416440665721893
it: 0.010309118777513504
here: 0.010284281335771084
and: 0.009794403798878193
and: 0.009078304283320904
it: 0.00885285809636116

Figure 2: attn tok-20词