

# 周报

屈原斌

2020 年 8 月 31 日

## 1 上周任务

- 相似度计算补充实验；
- 论文综述；
- Bert做多分类实验（实验结果有问题，正在修改）；

### 1.1 实验一

- 数据：初中作文，每个主题抽取10篇做范文，10篇做测试；
- 实验设置：
  - 每个主题随机从其他主题中抽取50篇作文构建负样本，计算prfa值作为评价标准；
  - 对于正样本，排名位置为top1则分类正确，否则分类错误；负样本则相反；
- 实验结果（见表1和表2）：
  - 字向量结果优于词向量；
  - 使用topk=3时的结果优于取平均值和最大值；
  - 使用output层表示的结果优于隐藏层；

表 1: 初中作文实验结果1（取隐藏层表示）

	Acc.	P	R	F1
字向量，取平均值	0.9625	0.9610	0.9625	0.9550
字向量，取最大值	0.9589	0.9615	0.9589	0.9521
字向量，取topk=3	0.9641	0.9665	0.9641	0.9584
词向量，取平均值	0.9599	0.9584	0.9599	0.9519
词向量，取最大值	0.9574	0.9602	0.9574	0.9508
词向量，取topk=3	0.9627	0.9639	0.9627	0.9567

表 2: 初中作文实验结果2（取output层表示）

	Acc.	P	R	F1
字向量，取平均值	0.9736	0.9747	0.9736	0.9709
字向量，取最大值	0.9711	0.9729	0.9711	0.9678
字向量，取topk=3	0.9742	0.9757	0.9742	0.9717
词向量，取平均值	0.9717	0.9730	0.9717	0.9687
词向量，取最大值	0.9710	0.9727	0.9710	0.9680
词向量，取topk=3	0.9720	0.9737	0.9720	0.9692

## 1.2 实验二

- 数据：小学作文，每个主题抽取10篇做范文，10篇做测试；
- 实验设置：
  - 每个主题随机从其他主题中抽取20篇作文构建负样本，计算prfa值作为评价标准；
  - 对于正样本，排名位置为top1则分类正确，否则分类错误；负样本则相反；
- 实验结果：
  - 四年级作文（见表3和表4）：
    - \* 字向量结果优于词向量；
    - \* 使用字向量时取topk=3的方法结果较差；
    - \* 使用output层表示的结果优于隐藏层；
  - 五年级作文（见表5和表6）：
    - \* 字向量结果优于词向量；
    - \* 使用topk=3的结果较取最大值和平均值的结果差；
    - \* 使用output层表示的结果较隐藏层表示结果差；

表 3: 小学作文（四年级）实验结果1（取隐藏层表示）

	Acc.	P	R	F1
字向量，取平均值	0.8400	0.8576	0.8400	0.8361
字向量，取最大值	0.8267	0.8492	0.8267	0.8218
字向量，取topk=3	0.8267	0.8379	0.8267	0.8205
词向量，取平均值	0.7533	0.8081	0.7533	0.7346
词向量，取最大值	0.7667	0.7955	0.7667	0.7371
词向量，取topk=3	0.8067	0.8399	0.8067	0.8003

表 4: 小学作文（四年级）实验结果2（取output层表示）

	Acc.	P	R	F1
字向量，取平均值	0.8400	0.8755	0.8400	0.8204
字向量，取最大值	0.8400	0.8521	0.8400	0.8347
字向量，取topk=3	0.8267	0.8444	0.8267	0.8222
词向量，取平均值	0.7733	0.7806	0.7733	0.7363
词向量，取最大值	0.7933	0.8063	0.7933	0.7771
词向量，取topk=3	0.8133	0.8516	0.8133	0.7917

表 5: 小学作文（五年级）实验结果1（取隐藏层表示）

	Acc.	P	R	F1
字向量，取平均值	0.8733	0.8893	0.8733	0.8720
字向量，取最大值	0.8267	0.8395	0.8267	0.8273
字向量，取topk=3	0.8533	0.8726	0.8533	0.8524
词向量，取平均值	0.8400	0.8629	0.8400	0.8360
词向量，取最大值	0.8133	0.8236	0.8133	0.8078
词向量，取topk=3	0.7933	0.8239	0.7933	0.7920

表 6: 小学作文（五年级）实验结果2（取output层表示）

	Acc.	P	R	F1
字向量，取平均值	0.8200	0.8668	0.8200	0.8125
字向量，取最大值	0.8400	0.8543	0.8400	0.8355
字向量，取topk=3	0.8267	0.8639	0.8267	0.8233
词向量，取平均值	0.7733	0.7876	0.7733	0.7687
词向量，取最大值	0.8133	0.8302	0.8133	0.8093
词向量，取topk=3	0.8200	0.8296	0.8200	0.8185

## 2 本周计划

- 完成Bert分类实验；
- 完成论文综述；