

周报

2020 年 4 月 20 日

1 上周任务

- 重新对非隐喻词进行分析；
- 看论文。

2 问题

- 之前的模型中没有cls参数，导致masked-lm模型预测不准确；

3 任务详细

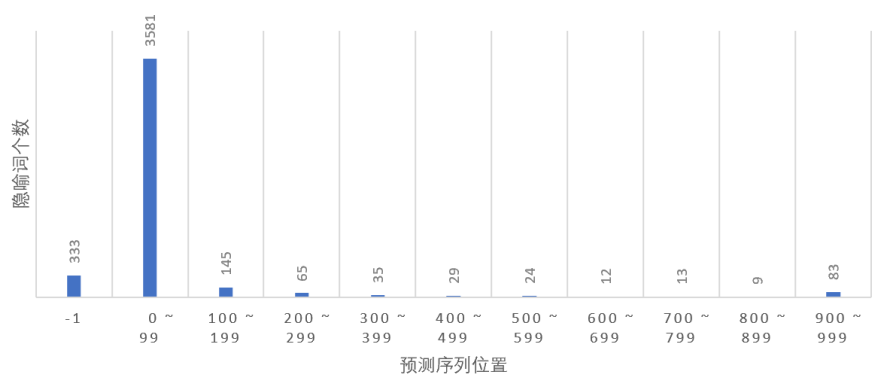
3.1 （非）隐喻词平均排名，见表1:

表 1: 平均排名

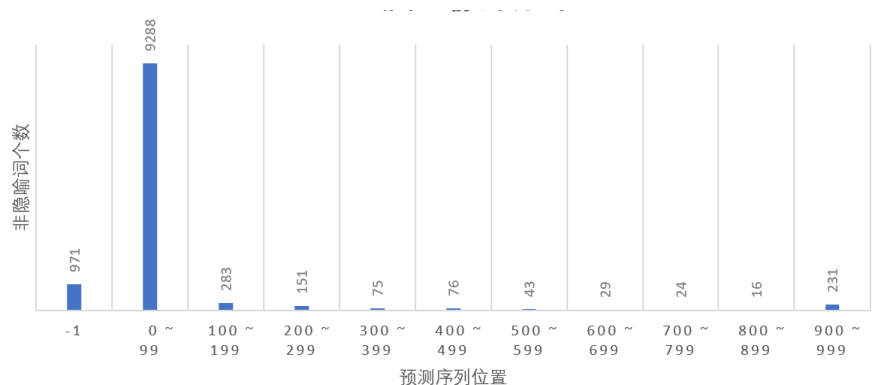
数据集	隐喻词		非隐喻词	
	个数	平均排名	个数	平均排名
VUA	4329	90	11187	116
TroFi	1627	220	2110	187
TroFi-X	601	147	843	60

3.2 （非）隐喻词具体分布，见图1、2、3：

- 基本分布在100以内，只统计了0-1000范围的分布情况；
- 绝大部分非隐喻词的位置比隐喻词位置靠前。



(a) 隐喻词分布

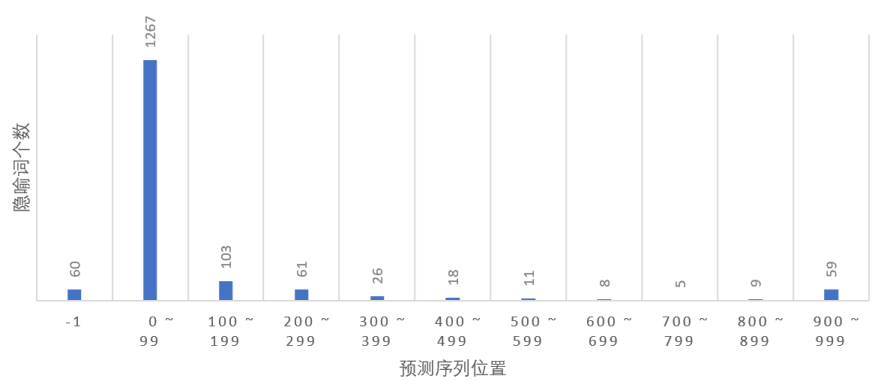


(b) 非隐喻词分布

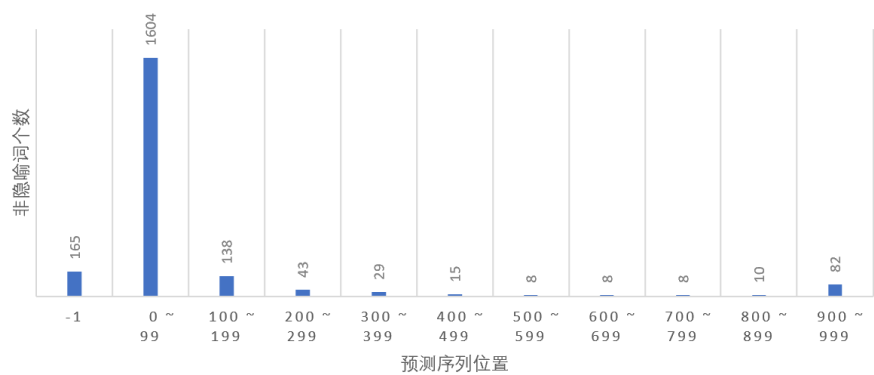
图 1: VUA Dataset

4 本周计划

- 复现一下郭师姐的代码；
- 看论文。

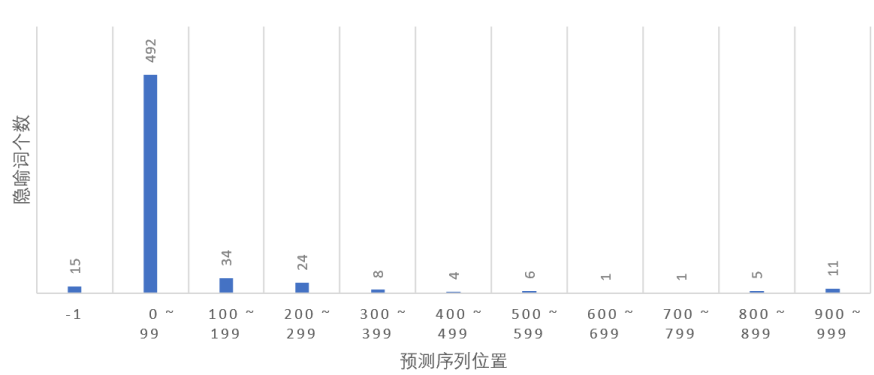


(a) 隐喻词分布

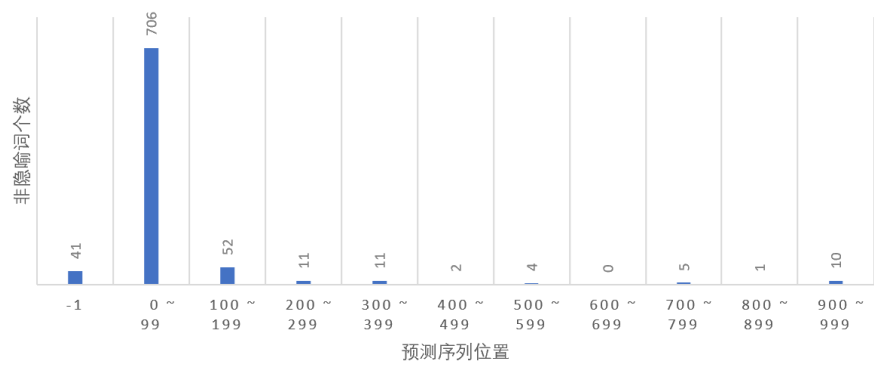


(b) 非隐喻词分布

图 2: TroFi Dataset



(a) 隐喻词分布



(b) 非隐喻词分布

图 3: TroFi-X Dataset