

北京大学 21-22 秋季学期《计算语言学》课程项目

概述

本次作业基于中文智源指数自然语言理解能力评测基准中的[古诗文识记数据集](#)。给定中国古典诗歌的现代文描述，同学们需要训练模型，使其能够从候选的四句诗中挑选出与给定的现代文描述意思符合的那一句诗歌。

数据集介绍

本次数据集基于中文智源指数自然语言理解能力评测基准中的[古诗文识记数据集](#)，提供带标注的训练集和验证集，验证集不提供标签，各个数据集的大小如下：

训练集：21,778 句；验证集：2,720 句；测试集（Unlabeled）：2,720 句。

样例

```
{"translation": "月夜朦胧，用碧玉做成的栏杆和用红色的砖砌成的墙是刺史的府宅。", "choices": ["兰玉当年刺史家", "碧砌红轩刺史家", "碧殿红棂翠浪间", "一簇楼台刺史家"], "answer": 1}
```

评价指标

使用分类准确率评价模型在**测试集**上的总体性能。

由于测试集标签未公开，请同学们将原始的验证集**均分**为验证集和测试集，根据自行切分的验证集调整参数，在自行切分的测试集上测试准确率。确认参数最优后，可以到[测评网站](#)进行评测。

参考文献

CCPM: A Chinese Classical Poetry Matching Dataset. Li, Wenhao and Qi, Fanchao and Sun, Maosong and Yi, Xiaoyuan and Zhang, Jiarui. arXiv preprint arXiv:2106.01979

实验报告

项目完成后，请各小组完成实验报告记录本次课程项目过程，内容大纲如下（可根据项目实际适当调整）：

1. 实验目的（阐述任务）
2. 实验原理（描述模型）
3. 实验内容（描述实验步骤，注重可复现性）
4. 实验结果（描述实验结果）
5. 实验分析（侧重对于结果和 case 的分析）
6. 实验过程总结和感想（每位组员都需要写）
7. 实验分工（写明各个组员的工作量）

其他

- 助教为大家提供了 baseline [实现](#)，仅供参考。
- 课程项目可以选择单人或者组队完成，每组不超过 3 人
- 项目作弊会被判定为 0 分，包括但不限于：实验造假、抄袭代码等

- 评分标准
 - 平时成绩：考勤
 - 大作业：报告 + 测评分数（自己切分测试集，和官方测评结果综合考虑）
- 禁止任何形式的在模型训练和提交结果中使用测试集标签信息，包括通过自己人工标注测试集标签
- 希望大家重视模型架构的可解释性，实验效果不是唯一评价标准，针对任务特点的改进，对于失败尝试的深入分析都是加分项
- 鼓励大家在项目上尝试一些创新的思路，包括但不限于模型的改动、任务形式的转换、从数据本身出发的思考等。