

2024 游戏安全技术竞赛比赛题目-复赛

(NLP&机器学习)

游戏跨语言恶意内容识别

赛程安排：



问题背景

游戏文本内容审核系统（Content Moderation System）旨在识别并检测游戏各 UGC（User Generated Content）场景中出现的违规文本，如色情低俗、侮辱谩骂等。在海外游戏环境中，游戏文本内容审核系统的建设面临多语言的挑战。参赛者需要基于英语数据集开发一款跨语言恶意文本识别模型，最终模型将在包含英语、阿语、土语、俄语的测试集上评估效果，并取各语种的 F-score 平均值作为评定最终成绩的依据。比赛不限制使用任何开源数据、模型、代码，只需要在最终提交文件中详细说明即可。

数据

1. 训练集及相关数据集：

- 8k 条带标注数据（英语），文件名：train.txt
- 4*20k 条无标注数据（每个语种各 20k），文件名：unlabel_text.txt
- 4*5k 条 ChatGPT 标注数据（每个语种各 5k），文件名：labeled_text_by_ChatGPT.txt。
prompt 见文件 labeled_text_by_ChatGPT_prompt.txt。
- 50k 平行语料（以英语为原语言，通过 ChatGPT 翻译获取）文件名：
parallel_text_by_ChatGPT.txt。prompt 见文件 parallel_text_by_ChatGPT_prompt.txt。

以上数据选手可自行分析，合理利用。

2. 验证集：

- 4*100 条带标注数据（每个语种各 100），文件名：dev_ar.txt、dev_en.txt、dev_ru.txt、dev_tr.txt

3. 测试集：4*1k 无标注数据（每个语种各 1k），与验证集同分布，不对外提供。

评估指标

采用每个语种测试集上的 $Fscore$ 值作为指标，如英语的指标为：

$$Fscore_{en} = (1 + \beta^2) \frac{precision_{en} * recall_{en}}{\beta^2 * precision_{en} + recall_{en}}$$

其中 $\beta = 0.7$ 。因为我们更加关注 $precision$ 。

最终评估指标为各语种 $Fscore$ 值的平均：

$$Fscore_{avg} = \frac{Fscore_{en} + Fscore_{ar} + Fscore_{tr} + Fscore_{ru}}{4}$$

上传文件说明

1. 请将所有相关文件打包为一个 zip 压缩包上传。
2. 压缩包命名方式“复赛-机器学习-姓名-学校-手机号”，如“复赛-机器学习-胡图图-翻斗大学-13333333333.zip”。
3. 压缩包中，必须包含以下文件：
 - a. requirements.txt，需写明所有第三方依赖，格式见“提交示例.zip”。
 - b. README.md，该文件中需详细写明解题思路，代码执行逻辑等。如果无法用纯文本描述，可附加其他常见格式文件。
 - c. [predict.sh](#)，执行预测的主脚本，详细信息见 4.
 - d. 其他预测脚本依赖代码、模型、文件。
4. 预测主脚本必须以“predict.sh”命名并放在项目的根目录下，我们最终将在项目根目录下执行“sh [predict.sh](#) {predict_file} {predict_result}”来获取预测结果。其中 predict_file 为待预测文件，格式为每一行一条文本“{文本}”，predict_result 为预测结果文件，格式必须为“{标签}{文本}”。两个文件的行需逐行对应，不得打乱顺序。示例如下：

```
0|*** you are
1|can only run with pan in hand
0|Only aov suck with q
1|Maiyen big nub
0|you are the jungler
1|bsdk c my play then tlk
0|Shtp meat
1|your bitch
0|Fck i lost this now
1|Dogs china no skills
```


注意事项

1. 允许使用外部资源，包括但不限于代码、工具、数据、模型，但要求所使用的资源是公开可获取的，并需要在提交的文件中详细给出名称、来源、数据量（如果是数据）、使用目的等全面的描述。
2. 我们最终将在 CPU 上执行预测，要求预测 QPS 达到 3，即每秒钟至少完成 3 条文本预测。如果 QPS 达不到要求，结果将作废。作为参考，twitter-xlm-roberta-base 模型在 CPU 上的 QPS 为 6.31。以下为执行预测的机器信息：

CPU：8 Intel(R) Xeon(R) Platinum 8255C CPU @ 2.50GHz

内存：16G

磁盘：150G

3. 请确保你的代码在以下操作系统下可以运行，如果有第三方依赖，请在依赖文件中写明。

```
No LSB modules are available.  
Distributor ID: Ubuntu  
Description:    Ubuntu 22.04.3 LTS  
Release:        22.04  
Codename:       jammy
```

BASELINE

采用 twitter-xlm-roberta-base 作为基础模型，<https://huggingface.co/cardiffnlp/twitter-xlm-roberta-base>

合并训练集和验证集作为训练集，训练参数：learning_rate=1e-5;batch_size=64;train_epoch=3;

最终指标：0.674094