



计算机科学
与技术系

实验三 嵌套命名实体识别

老师：戴新宇

助教：欧阳亚文，龙思宇

{ouyangyw, longsy}@smail.nju.edu.cn

或者在课程 QQ 群中联系

概要

- 实验任务
- 实验评分
- 实验提交
- 注意事项

实验任务

- 任务：嵌套命名实体识别
- 说明：命名实体识别(NER)是NLP中的基础任务，旨在抽取出句中的实体，并把它们分类为正确的实体类别。嵌套命名实体识别更为复杂，因为同时会考虑实体之间的嵌套。比如“南京大学”不仅是组织，“南京”同时是地点。
- 要求：给定一段文本，输出实体位置与类型，不同实体用'|'隔开。
 - 举例来说，文本是“南京大学有三个校区”，输出应是“0, 2 地点|0, 4 组织”。

实验任务

- 评估：f1
 - 上页的例子中，如果模型输出是“0, 2 地点 | 0, 3 组织”，精度是0.5，召回是0.5，f1是0.5。
 - 在评测时不会考虑实体顺序，如果模型输出是“0, 4 组织 | 0, 2 地点”，f1是1。
- 数据：通过QQ群发送给大家，包括train, dev, test, 大家需要提交test集抽取结果，**每个样本一行，没有实体则该行为空。**
- 要求：不能使用预训练语言模型，比如BERT。

实验评分

- 实验总分为100分
- 性能（**70%**）
 - 采用打榜的方式。
 - 达到bonus阈值获得满分。
 - 后面的会根据排名和不同排名之间分数的差距递减。
- 实验报告及代码风格、实现方式等（**30%**）
- 选做部分（bonus 30%）
 - 超过预先设定的SOTA结果（30%）
 - 超过阈值的前十名获得完整bonus
 - 超过bonus阈值越多给分越高

实验提交

- 排行榜提交: <http://114.212.189.62:12345/>
- 代码和报告提交: <http://cslabcms.nju.edu.cn>
 - 使用个人账号登录, 独立完成
 - 提交**第十四周**的 Project 3嵌套命名实体项目
 - 所有内容打包并压缩, 命名为学号.zip/rar/tar.gz, 如 181220001.zip, 重复提交需要先删除旧版本
 - 提交后请确认作业状态为“已经提交”、“已提交等待评分”, 而不是“草稿(未提交)”
- 上述过程中的任何问题请联系助教或老师, 无特殊情况不接受其他提交方式, 截止日期: **2021年 1月 10日, 23:55:59**, 请尽量不要在此之前的几分钟提交, 网络有风险。

代码和报告提交

- 压缩包内容至少包含
 - 源代码（推荐使用Python）
 - 尽量遵循优秀代码规范，有必要注释
 - 实验报告
 - 推荐 pdf， docx 也可以
 - 报告内容包括个人信息，实现的方法，运行方式，实验总结等
 - 不要贴大段代码，篇幅不超过 4 页
 - 使用的额外数据，无需提交训练测试数据集
 - README 可选

注意事项

- 若无法复现榜单结果的将**取消成绩**，所以如果你使用了随机算法，请固定随机种子
- 请不要将任何开源项目/工具包的运行结果作为提交文件，一旦发现将**取消成绩**
- 本次实验**不允许**使用额外数据，若对使用的其他资源是否属于额外数据有疑问的，请及时咨询助教
- 参考网上的任何代码请指明出处！区别参考与抄袭，任何形式的代码抄袭都是不允许的！被确认的抄袭者与抄袭者本次实验都**记为 0 分**！
- 代码中不要出现绝对路径
- 运行方式最好写成指令的形式

Thank you