

## 数智NLP组2022 Final Assessment

(4 weeks)

Introduction

Final Assessment

Basic Library

Platform

task

Requirement

Text pre-processing (25)

Use tensorflow(20)

Requirements for document (55)

GUI(20)

Respondent (100)

课内重要!!! 不要挂科!!!

# 数智NLP组2022 Final Assessment

## (4 weeks)

## Introduction

深度学习，是一个非常需要花时间进行研究方向。不考研或者中途想要转方向的师弟师妹们，我在此劝说一下：不要把争取来的机会随意丢弃；不要随风飘动改变自己的志向。虽然科研不一定能成功，但贪松怕卷勿入斯门，心藏开发请往他处。

这次不说那么多啦！冲起来，就快结束了，就快成功了！！！！

## Final Assessment

### Basic Library

- tensorflow2.x 具体不限制。

### Platform

- Colab

### task

接下来就是正经活了！这是最简单的任务了：**文本分类**：来自yelp的数据集。

### Requirement

#### Text pre-processing (25)

- 自行构建词表
- 英文使用Subword算法：**BPE**（不是用框架能够额外+5）也可以使用其他算法。
- 是否需要进行Embedding？
- 考虑设备问题，公平起见 每一个word\_dim<=128
- 公平起见 sequence 长度不能超过100

## Usa tensorflow(20)

1. **允许**使用**tensorflow**基础框架【禁止使用别人搭建好的代码】
  1. 允许使用CNN、LSTM等模块
  2. 公平起见: 要求参数不超过4M, 可训练参数不超过1M
  3. 使用必须在建立在学习过的条件下, 并完成使用文档撰写 (越详尽越好)
2. **全连接层(Dense)继承keras.layers.Layer来自己写, 不直接使用Dense, 自定义其为Linear类 (也就是要求: class Linear(这里应该是什么我就不提示了) 不懂的话搜索一下keras官方文档) (额外+5)**
3. 在防止过拟合方面可使用框架对应部分, 不需要手写, 文档中要求使用的部分要有体现【可以加粗字体等强调形式】 (额外+10)
4. 损失函数的选择恰当
5. 在进行实验的过程中请记录过程。 (额外+3)

## Requirements for document (55)

跟上次一样, 往详细、清晰、干净的学术论文的感觉去靠.

## GUI(20)

## Respondent (100)

提交文档格式: 如果不对, 我看心情扣分哦~~

```
Submission_name/  
...data/  
.....data.csv  
.....etc.  
...code  
...doc  
etc.
```

**课内重要!!! 不要挂科!!!**