

# Lab 2

---

By ZYT 2022/4/10 V3

## 一、实验要求

编写RNN的语言模型，并基于训练好的词向量，编写RNN模型用于文本分类。

本次实验内容

- 实现SimpleRNN, LSTM, GRU(基于训练好的词向量Glove)，并对结果进行对比
- 对比单层RNN与两层RNN
- 对比隐藏神经元数量对结果的影响

## 二、数据集

本次实验我们采用IMDB数据集

<http://ai.stanford.edu/~amaas/data/sentiment/>

这是一个多分类情感分类的数据集，包含了5万条流行电影的评论数据，其中训练集25000条，测试集25000条，本次实验只要求做二分类任务

## 三、Glove

原理:[GloVe原理详解](#)

使用:

- Pytorch [https://blog.csdn.net/qg\\_40711769/article/details/105416442](https://blog.csdn.net/qg_40711769/article/details/105416442)
- TF <https://blog.csdn.net/bqw18744018044/article/details/89575127>

## 三、代码框架

本次实验提供Pytorch版本的RNN+Glove代码框架，数据已处理好，你需要填充model以及training部分

如果因环境问题无法使用 `torchtext`，可参考[pytorch](#)

TF可参考[gloVe预训练-imdb数据集-Keras](#)

同上次实验，你完全可以不参考我提供的任何东西~

## 四、提交要求

你需要提交

- 代码、报告、ReadMe
- 请不要把数据放在压缩包里

压缩包命名 PB18XXX\_姓名\_Lab2.zip

在**2022.4.24 23:59:59** 之前提交到[deeplearning2022s@163.com](mailto:deeplearning2022s@163.com)

友情提示各位DDL战士至少提前三天开始^v^

## Update V2

- 应同学们要求，新增无需 `torchtext` 的数据加载形式，但需要你掌握 `nn.Embedding`
  - 如果你采用这种方式，仅需要看 `Lab2_New` 文件
  - 你需要按照 `ReadMe` 的文件结构放置数据集以及 `Glove` (你可以更改代码中的路径)
- 同时，考虑到 `torchtext` 更为简便一点，我提供了一个 `torchtext` 的安装环境
  - 如果你采用这种方式，仅需要看 `Lab2_old` 文件
  - 你需要访问来下载 `pip_packages` 文件夹并按照 `ReadMe` 的文件结构放置

链接: <https://rec.ustc.edu.cn/share/d2cdfb00-b82c-11ec-8fa0-772daf4e8153> 密码: 4qy8

- 在 `Lab2_old` 根目录下执行
  1. `conda create --name Lab2 python=3.6.13`
  2. `conda activate Lab2`
  3. `pip install --no-index --find-links=./pip_packages -r requirements.txt`
  4. 找到环境根目录下的 `torchtext`，如我的就是  
`D:\software\anaconda3\envs\test\lib\site-packages\torchtext\data\utils.py`，将 `utils.py` 的114行改为 `spacy = spacy.load("en_core_web_sm")`
  5. 至此，你就可以快乐的编写代码啦

## Update V3

更改 `requirements.txt` 中的绝对路径