20210522-周报总结

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **小论文进展：** | **🞎阅读文献** | **🞎做实验** | **🞎撰写中** | **🗹其它** |
| **一、本周工作**   1. 稍微优化了一下，之前手写体的RNN，之前只有0.6左右的正确率，现在又0.8以上。 2. RNN实例LSTM和GRU，One-Hot编码（独热编码），jieba分词。 3. Python爬虫实例，bs和正则。 | | | | |
| 1. **下周计划** 2. 基础LSTM 做语言情感分类，积极和消极。 | | | | |
| 1. **本周工作摘要**   **问题一**：什么是LSTM和GRU？  **解决一**：  *LSTM：*  RNN的缺点：梯度消失、梯度爆炸；rnn较其他cnn和全连接要用更多的显存空间，更难训练；如果采用tanh、relu为激活函数，没法处理太长的序列  **为了解决梯度消失和爆炸**以及更好**的预测和分类**序列数据等问题，rnn逐渐转变为lstm    *GRU:*  因为LSTM的训练比较慢，而GRU在其上稍微修改，速度可以快很多，而精度基本不变。  自然语言处理NLP  一个句子：如何使用张量表示一个句子  不同的句子，长度是不一样的   问题二：OneHot编码：独热编码 我爱中国 词表：4个不同的字 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 每一个字使用 4个特征组成，只有一个位置是1，其它位置都是0 问题1：稀疏矩阵，如果词表比较大，则会占据很大的空间 问题2：相关性为0      每个单词构成一个向量  ## 词嵌入  n\_vacab = 8  emb = nn.Embedding(n\_vacab, 4) # 学习weight  X = emb(dd)  print(X.shape) # (batch, 输入特征, 输出特征)  ### N个句子，每个句子M个字，每个字有X个特征 | | | | |