

基础题答案:

<https://www.jianshu.com/p/04225bc9c8f8>

<https://www.jianshu.com/p/8f19b04fc0eb>

附加题:

BI:

2.你在使用deepFM的时候是如何处理欠拟合和过拟合问题的 I

欠拟合: 增加deep部分的层数, 增加epoch的轮数, 增加learning rate, 减少正则化力度 I 过

拟合: 在deep层直接增加dropout的率, 减少epoch轮数, 增加更多的数据, 增加正则化力度, shuffle数据

NLP:

BERT和GPT的异同点

解析:

1.GPT在BooksCorpus(800M单词)训练; BERT在BooksCorpus(800M单词)和维基百科(2,500M单词)训练。

2.GPT使用一种句子分隔符([SEP])和分类符词块([CLS]), 它们仅在微调时引入; BERT在预训练期间学习[SEP], [CLS]和句子A/B嵌入。

3.GPT用一个批量32,000单词训练1M步; BERT用一个批量128,000单词训练1M步。

4.GPT对所有微调实验使用的 $5e-5$ 相同学习率; BERT选择特定于任务的微调学习率, 在开发集表现最佳。

CV:

LSTM算法的不足之处有哪些?

2.2 缺点总结

(1) RNN的梯度问题在LSTM及其变种里面得到了一定程度的解决, 但还是不够。它可以处理100个量级的序列, 而对于1000个量级, 或者更长的序列则依然会显得很棘手。

(2) 计算费时。每一个LSTM的cell里面都意味着有4个全连接层(MLP), 如果LSTM的时间跨度很大, 并且网络又很深, 这个计算量会很大, 很耗时。
