深度之眼
deepshare.net

学员问题解答

对大家提出的问题进行解答

扫码咨询客服

本次答疑内容

Course content

1、重难点知识串讲

2、学员问题解答

3、互动答疑

重难点知识串讲

trick

1. 调参，lr | batch_size | dropout | bert最后一层，还是最后多层 | bert+cnn？抑或bert+rnn？

2. 提高recall，从目前线上来看，召回偏低，对于解码部分比较严格，会丢失一部分预测结果，可想办法尽可能控制准确率提高召回。如果没有好的思路，可采用模型融合的方式进行召回补充。

3. 构造训练样本方式上，由于训练样本很多偏长（大于512），可以尝试CNN卷积划窗的形式。以某个特殊符号进行切分，设定窗口大小。

重难点知识串讲

trick

4. 模型融合（很重要的上分点）| PS：多实验，多记录过程，实时保存最好的模型文件，最后进行模型融合。

a) 概率融合； b） 投票融合。

5. 半监督迁移学习【 网上找公开的医学相关数据，最好数据分布差异小 】

6. 使用chinese_roberta_wwm_large_ext_L-24_H-1024_A-16（对机器要求高）

重难点知识串讲

Semi-Supervised DA

深度之眼
deepshare.net

重点

$$L = min\left(\frac{1}{n}\sum_{i=1}^{n} W_1 * L(\theta\ (x_i)\ , y_i, \theta) + \frac{1}{m}\sum_{i=1}^{m} W_2 * L(\theta\ (x_i)\ , y_i, \theta) \right)$$

$$= min\frac{1}{n+m}\sum_{i=1}^{n+m} W_i * L(\theta\ (x_i)\ , y_i, \theta)$$

其中Wi，是每个样本的权重。有两种设计思路：

1.固定W1与W2值，经验值,真实标签个数为n，未标注样本个数为m，W1对应真实样本权重，W2对应未标注样本权重。

经验值：m/n 在 [5*W1/W2,10*W1/W2]

2.探索一下线上与线下的标签分布，根据标签分布进行W权重的调整（扰动）。

答疑问题

Answer questions

深度之眼
deepshare.net

疑问1：我安装tensorflow2.1之后运行baseline代码出现图中的错误，这个placeholder我记的是1.x才有，老师给的baseline不应该是基于2.1的嘛，为什么会报着个错啊

```
    539         x = tf.sparse_placeholder(dtype, shape=shape, name=name
    540     else:
--> 541         x = tf.placeholder(dtype, shape=shape, name=name)
    542     x._keras_shape = shape
    543     x._uses_learning_phase = False

AttributeError: module 'tensorflow' has no attribute 'placeholder'
```

```
]:
class NamedEntityRecognizer(ViterbiDecoder):
    """命名实体识别器
    """
    def recognize(self, text):
```

答疑问题

Answer questions

疑问2：请问pytorch在将梯度张量转numpy的时候，使用Tensor.data.numpy()和Tensor.detach().numpy()的区别是什么

x.data和x.detach()新分离出来的tensor的requires_grad=False，即不可求导时两者之间没有区别，但是当requires_grad=True的时候的两者之间的是有不同：x.data不能被autograd追踪求微分，但是x.detach可以被autograd()追踪求导。

答疑问题

Answer questions

深度之眼
deepshare.net

```
b = torch.tensor([1,2,3.], requires_grad=True)
out = b.sigmoid()
out
-----------------------------------------------
output:tensor([0.7311, 0.8808, 0.9526], grad_fn=<SigmoidBackward>)

c = out.detach()
c
-----------------------------------------------
output:tensor([0.7311, 0.8808, 0.9526])

c.zero_()       # # out的值被c.zero_()修改 !!
out.sum().backward()   # 报错是是因为autograd追踪求导的时候发现数据已经发生改变，被覆盖。
-----------------------------------------------
output: RuntimeError: one of the variables needed for gradient computation has been modified by an
inplace operation:
```

答疑问题

Answer questions

深度之眼
deepshare.net

疑问3：深度学习调参，一般是手动调参，还是自动调参哈

答疑问题

Answer questions

深度之眼
deepshare.net

疑问4：目前模型的召回率比较低，我考虑的方法是否可以在解码阶段优化一下，如果模型返回的top3或者top5的实体都加进来，这样虽然概率比较低，但是召回无疑都提升了。然后相比给这些结果打分，或者模型融合。老师看看是否可行？

答疑问题

Answer questions

疑问5： 1. 4-5个epoch直接开始过拟合，如何减少过拟合？

2. 如何使用中间层进行融合输出

3. 使用bert4keras是否属于使用过多开源代码？

- **Google原版bert**: https://github.com/google-research/bert
- **brightmart版roberta**: https://github.com/brightmart/roberta_zh
- **哈工大版roberta**: https://github.com/ymcui/Chinese-BERT-wwm
- **Google原版albert**[例子]: https://github.com/google-research/ALBERT
- **brightmart版albert**: https://github.com/brightmart/albert_zh
- **转换后的albert**: https://github.com/bojone/albert_zh
- **华为的NEZHA**: https://github.com/huawei-noah/Pretrained-Language-Model/tree/master/NEZHA-TensorFlow
- **华为的NEZHA-GEN**: https://github.com/huawei-noah/Pretrained-Language-Model/tree/master/NEZHA-Gen-TensorFlow
- **自研语言模型**: https://github.com/ZhuiyiTechnology/pretrained-models
- **T5模型**: https://github.com/google-research/text-to-text-transfer-transformer
- **GPT_OpenAI**: https://github.com/bojone/CDial-GPT-tf
- **GPT2_ML**: https://github.com/imcaspar/gpt2-ml
- **Google原版ELECTRA**: https://github.com/google-research/electra
- **哈工大版ELECTRA**: https://github.com/ymcui/Chinese-ELECTRA
- **CLUE版ELECTRA**: https://github.com/CLUEbenchmark/ELECTRA
- **LaBSE（多国语言BERT）**: https://github.com/bojone/labse
- **Chinese-GEN项目下的模型**: https://github.com/bojone/chinese-gen

答疑问题

Answer questions

疑问6： bert4keras加载预训练模型后，可以用新的数据集继续预训练吗？可以的话，大概怎么做呢？希望老师能稍微详细地讲一下，bert4keras的文档真的太少了。

- 加载bert/roberta/albert的预训练权重进行finetune；
- 实现语言模型、seq2seq所需要的attention mask；
- 丰富的examples；
- 从零预训练代码（支持TPU、多GPU，请看pretraining）；
- 兼容keras、tf.keras

答疑问题

Answer questions

深度之眼
deepshare.net

疑问7： 老师好，请问下您说的抽取BERT的多层线性加权的模型融合中，是否仅是BERT的每个Block中feednorm层的输出，还是其它种类层的输出也行。

深度之眼
deepshare.net

互动时间

扫码咨询客服

AI比赛年度会员

Kuai lai jia ru us!

深度之眼
deepshare.net

Step0：选修知识

- 数学基础
- Python基础
- 图像基础
- NLP基础

深度方向

解决**基础不牢固**
替你**查漏补缺**

Step1：参加经典赛练习

四大方向+九场经典赛

- 数据科学
- NLP方向
- CV方向
- 综合方向

按照个人学习能力和技术深度，设计了不同阶段课程，带你**层层提升**。

https://ai.deepshare.net/all/3279059

Step2：参加进行的新比赛

Kaggle　DC竞赛 www.dcjingsai.com

TIANCHI天池

DataFountain　Kesci

轻松入门CV /NLP
扎实**细分领域**

添加小享回复【阿水】
获得比赛会员优惠券→
优惠仅限今晚！

Step3：上TOP

拿奖金

奖励/内推/实习

PS欢迎来当讲师
（长期跪舔TOP大神）

结语

我 说

感谢同学们参加今晚的直播答疑！

课下，请好好**总结和回顾知识点**

深度之眼
deepshare.net

联系我们：

电话：18001992849

邮箱：service@deepshare.net

公众号

客服微信