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

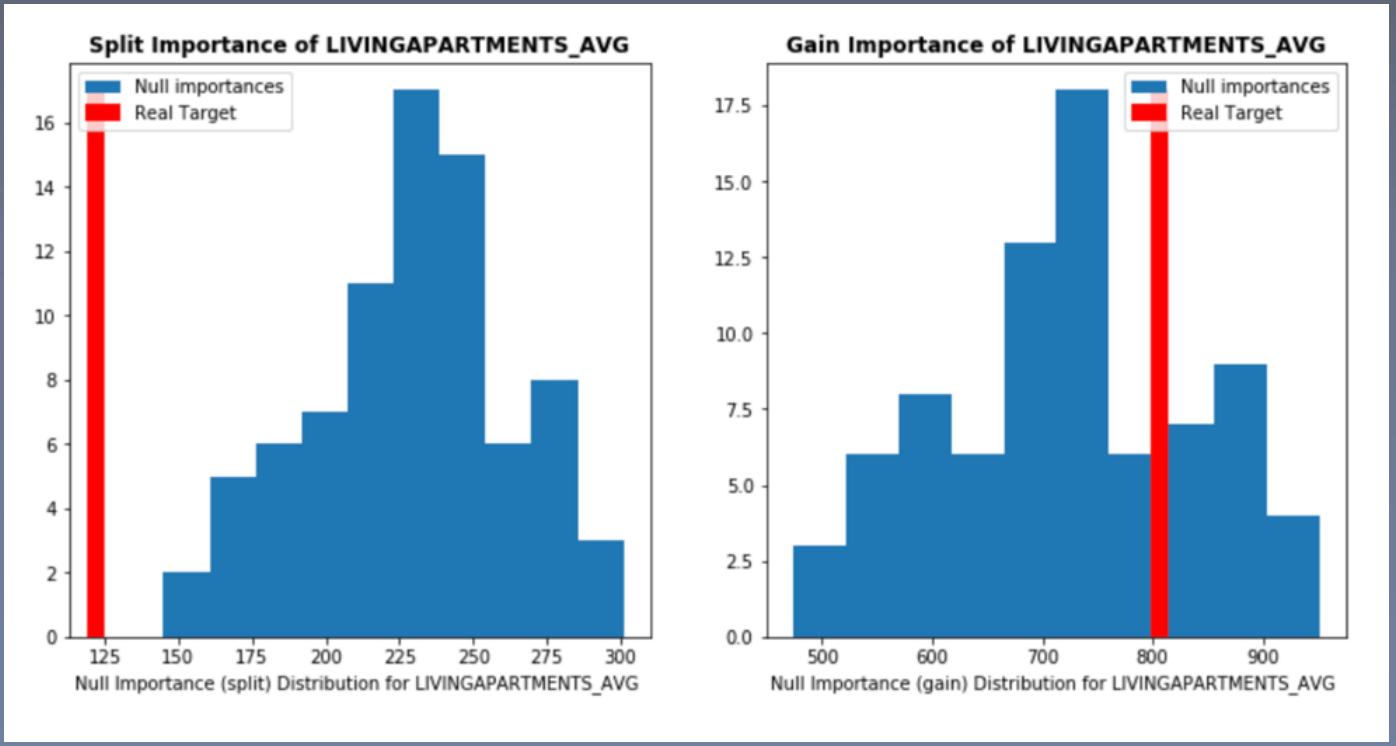
2、学员问题解答

3、互动答疑

答疑问题

Answer questions

疑问1： 这个null Importances 的图怎么解读呢?

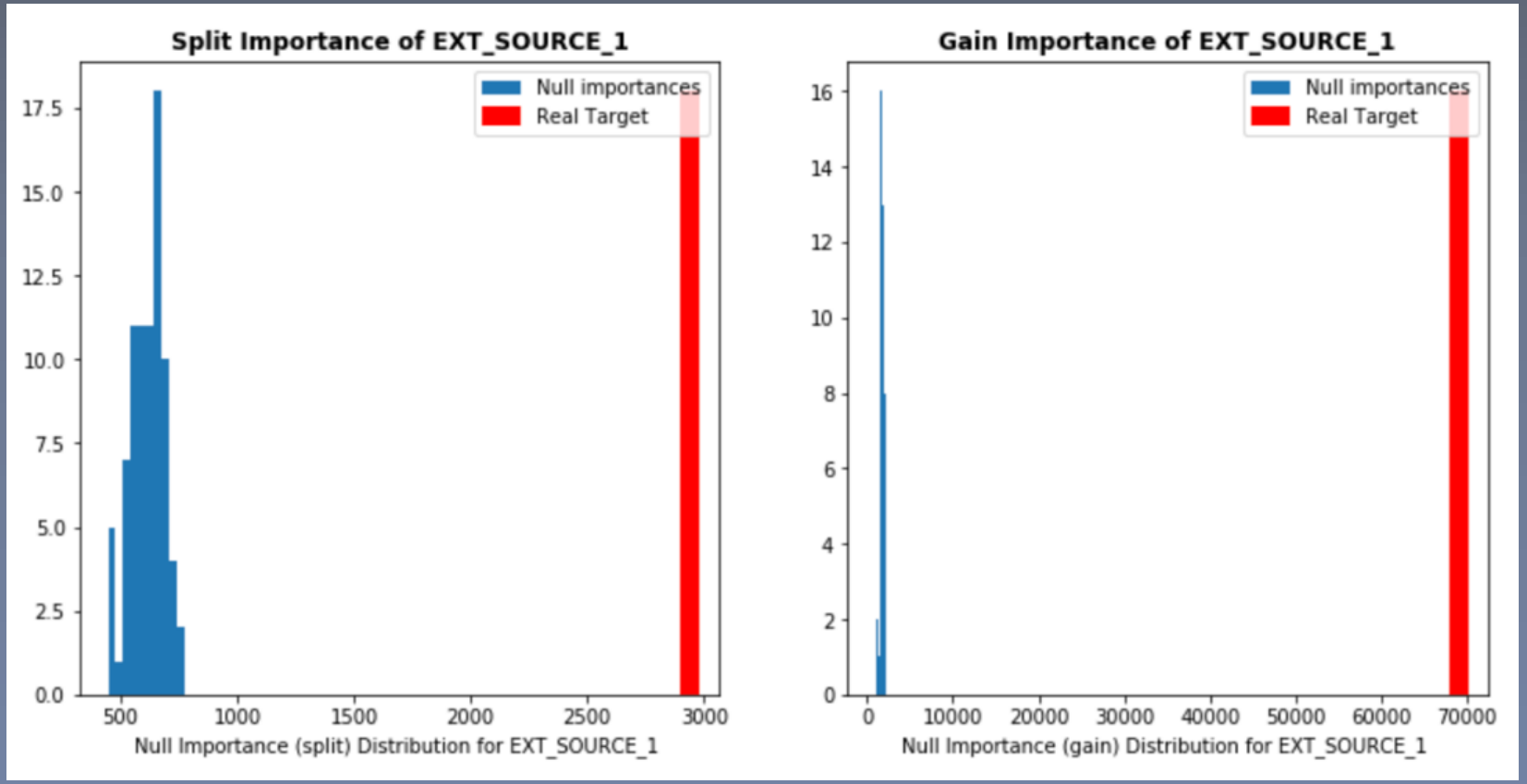


https://www.kaggle.com/ogrellier/feature-selection-with-null-importances

答疑问题

Answer questions

疑问2： 下面的x坐标是什么意思呢？ 如何区别特征好坏



答疑问题

Answer questions

疑问3：关于本阶段学习的问题： 个人觉得是有那么多集成学习的手段，想看老师具体演示一次有各种模型怎么筛选的coding实验过程。包括简单的ensemble求平均投票或者stacking

1、特征筛选

https://www.kaggle.com/arthurtok/introduction-to-ensembling-stacking-in-python

答疑问题

Answer questions

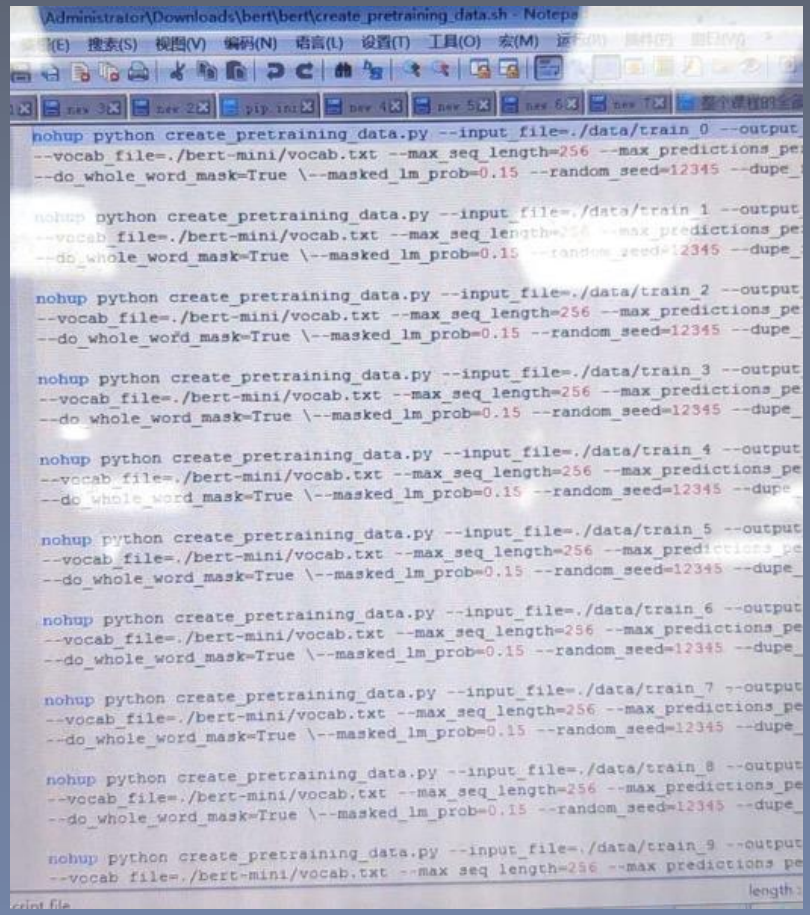
疑问4：老师能讲下bert预训练部分的代码吗？还有对于长文本，如果不想对文本截断，还有什么好方式吗？

1、把训练集和测试集数据整合在一起；

2、生成tfrecord文件；

3、进行pertrain；

4、如果不截断，maxlen越大越好；

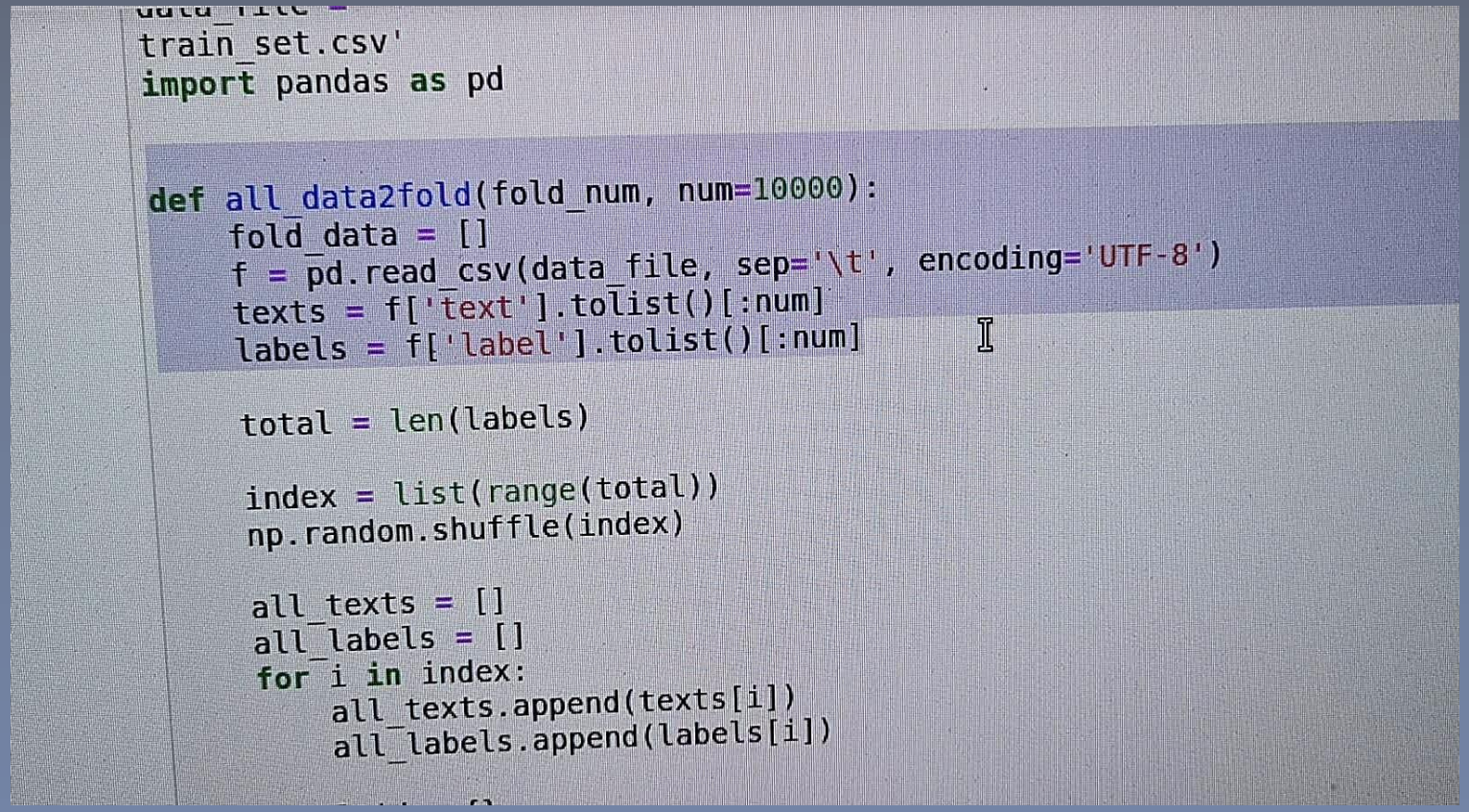


答疑问题

Answer questions

深度之眼
deepshare.net

疑问5：小白跑了dbc天池环境里的textcnn，只去掉了这里3个num，结果好像不是对测试集的预测 跑bert也遇到这个问题了



```
train_set.csv'
import pandas as pd


def all_data2fold(fold_num, num=10000):
    fold_data = []
    f = pd.read_csv(data_file, sep='\t', encoding='UTF-8')
    texts = f['text'].tolist()[:num]
    labels = f['label'].tolist()[:num]

    total = len(labels)

    index = list(range(total))
    np.random.shuffle(index)

    all_texts = []
    all_labels = []
    for i in index:
        all_texts.append(texts[i])
        all_labels.append(labels[i])
```

答疑问题

Answer questions

疑问6： 如何快速调参如learning rate，warm up steps的超参数；

1、理解超参数含义；

2、经验知识 + 历史总结；

答疑问题

Answer questions

疑问7：如何调整模型结构使得模型对超参数没那么敏感？

答疑问题

Answer questions

疑问8：当loss下降很慢时，如何判断应该优化参数或者模型结构，还是静心等待模型跑两三个小时出结果？

最好的方法：一台机器做继续优化（100 epoch），一台机器跑对比实验（20 epoch）；

答疑问题

Answer questions

疑问9： GPU使用率很高了70%+，但代码总是跑的很慢怎么办？

代码速度取决于：GPU速度、CPU速度、内存速度、磁盘速度；

答疑问题

Answer questions

深度之眼
deepshare.net

疑问10：有必要自己配一台GPU机器用来学习么？

自己配成本：一个2080ti + 32G内存，1w人民币，能用5年（每个月电费5百）；

https://developer.nvidia.com/cuda-gpus

深度之眼
deepshare.net

互动时间

扫码咨询客服

互动时间

Answer questions

互动1： NLP如何准备面试?

1、各种深度学习模型，如：CNN、RNN、LSTM、GRU、seq2seq、attention机制等。

2、各种机器学习算法：如logistic regression，decision tree, SVM,HMM,CRF,包括原理、公式推导、应用、优缺点等

3、各种词向量离散表示，例如BOW,TF-IDF,N-gram，分布式表示Word2Vec

4、NLP领域当前热点和前沿技术，如elmo、Transformer、Bert、Gpt、xlnet、Roberta等

互动时间

Answer questions

互动2：NLP学习方向如何选择?

1、掌握文本分类是基础，不要局限在文本分类；

2、建议结合其他任务一起学习：文本分类、语义理解、知识图谱构建、篇章理解、情感分析、自然语言生成；

互动时间

Answer questions

互动3： NLP就业方向如何选择?

1、掌握NLP每个任务落地和应用，找对应的岗位；

2、找准从业领域和交叉方向，搜索、推荐系统中的NLP；

深度之眼
deepshare.net

互动时间

Answer questions

互动4：下一场比赛学习什么？

1、图像检索走起！

2、图像检索的特征提取也是一种embedding；

赛题：数码设备图像检测

精确的图像检索是拍照购场景下的核心技术，是学术界和工业界的研究热点。

本题目将向选手提供真实拍照购场景下用于数码设备图像检索的训练数据集，包含手机、智能穿戴、PC、平板、音箱、路由器等粗粒度数码设备图像和细粒度（如不同型号外观手机等）数码设备图像，完成数码设备图像检索任务，即给定一张罕有数码设备的查询图像，算法需要在数码设备库中查找并返回罕有该商品的图像。比赛选手基于给定的训练数据构建并提交模型，比赛使用top-1检索准确率和mAP@10来评价比赛结果。

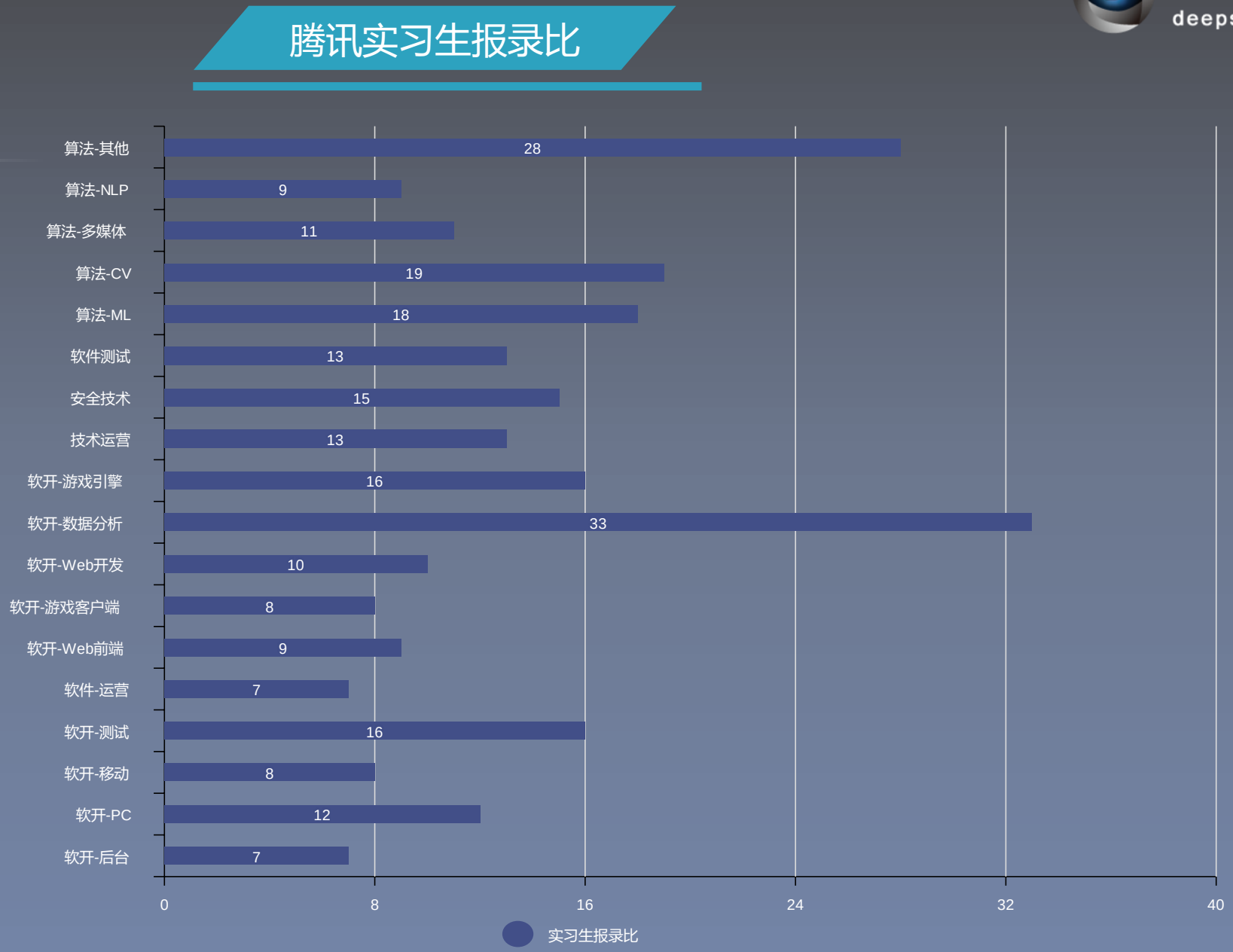
希望通过本次大赛发现图像检索算法领域的人才，推动该领域的发展。

互动时间

Answer questions

互动5： NLP就业机会

深度之眼
deepshare.net



互动时间

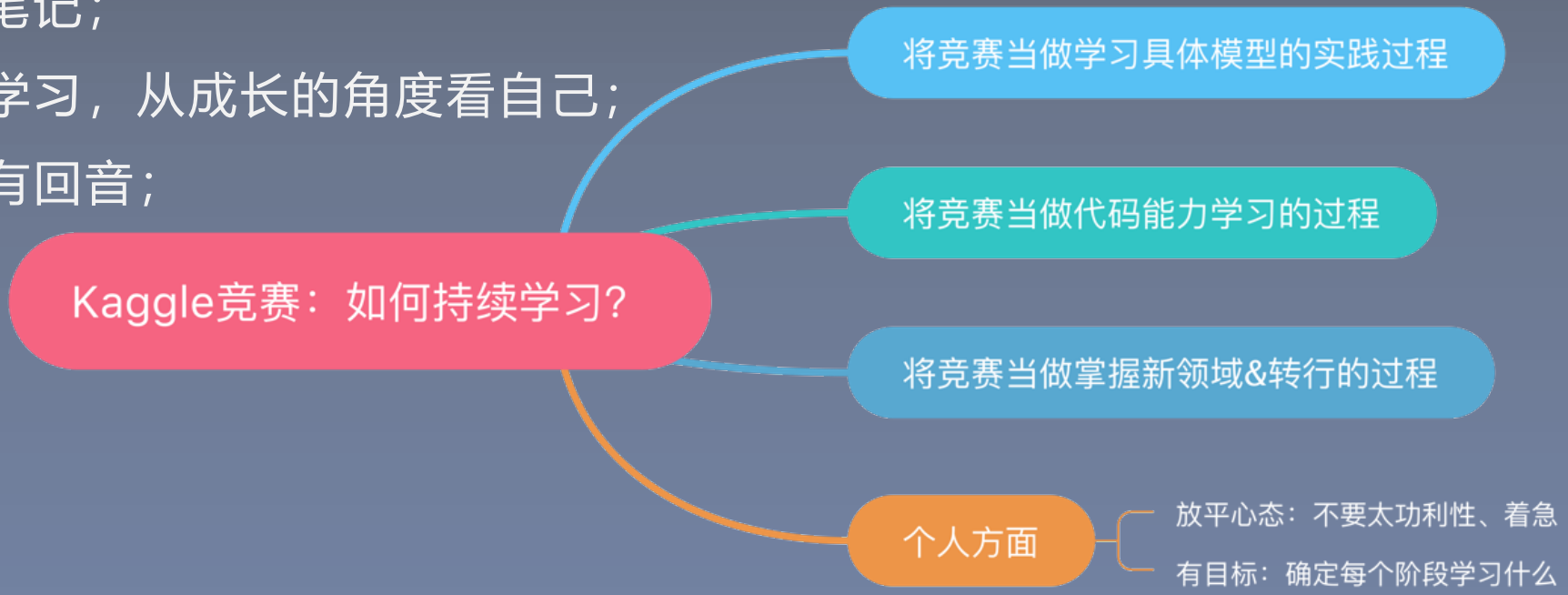
Answer questions

互动6：如何持续学习

1、降低期待感，做好笔记；

2、有连续型和持续的学习，从成长的角度看自己；

3、学习有计划，事事有回音；



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**。

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)

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

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