

Kaggle NFL头盔跟踪 首发第一堂直播课

导师：Mozak



更多干货，关注公众号 扫码获取比赛baseline

个人介绍

All about me

Mozak，人邮出版社签约作者

天池数据大神，Kaggle排名**前1000**；

四维图新自动驾驶算法赛**冠军**；

多次在国内外**数据竞赛**中获得优异成绩；

在深度之眼累计指导20+学生，获得银盘。

知乎

Kaggle



TIANCHI天池

个人介绍

All about me

CV竞赛对我个人：

- ✓ 改变了学习的方式：从知识点学习转变为从问题开始学习；
- ✓ 改变了生活的方式：个人变得积极主动，勇于尝试；
- ✓ 扩大了学习的范围：从结构化比赛到CV比赛，知识点++；

思考：结构化比赛与CV赛的主要区别是什么？



添加汪汪酱
获取大赛Baseline

目录

1/ 数据竞赛介绍 (10min)

2/ 赛题介绍 (15min)

3/ 赛题思路 (15min)

4/ 代码讲解 (15min)

5/ 经验分享 (5min)



更多干货，关注公众号 扫码获取比赛baseline

1 入门数据竞赛

from 0 to 1

Part1 数据竞赛是什么？

What is Data Competition ?

数据竞赛是以工业/学术**问题为导向**的，聚合广泛、跨学科人才参与的，利用数据研发算法模型和探索解决方案的研发模式。

数据竞赛是：

- ✓ 一种众包的竞赛模型（对参赛人员门槛没有限制）；
- ✓ 有明确的问题背景（具有较强的业务或数据背景）；

	数据竞赛	普通学科竞赛
内容	数据挖掘 & 机器学习	学科知识点
打分反馈	支持	不支持
定量打分	支持	不支持
交互式	支持	不支持

Part1 数据竞赛是什么？

What is Data Competition ?

■ 数据竞赛平台：Kaggle/天池/DC竞赛/DataFountain/FlyAI等

Kaggle是全球最大的数据竞赛平台，每年会举办几十场竞赛，主要以算法赛和可视化比赛居多。Kaggle具有完整的比赛平台机制：从赛题介绍、数据分析、评分、排名和最终的分享。

TIANCHI天池

Kaggle

DC竞赛
www.dcjingsai.com

Kesci

DataFountain

FlyAI

Baidu 百度 | AI Studio

讯飞开放平台
OPEN PLATFORM

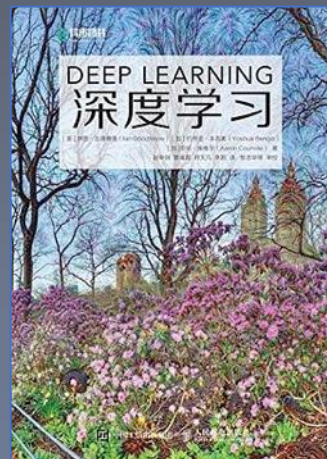
Part1 数据竞赛是什么？

What is Data Competition ?

■ 参加数据竞赛的意义

与理论相比，动手能力也同样重要：

- ✓ 对数据挖掘流程的了解；
- ✓ 对模型调参的了解；
- ✓ 对工具库的熟练程度；



1.1 数据竞赛是什么？

What is Data Competition ?

CV赛题占据40%以上!



Understanding Clouds from Satellite Images

Can you classify cloud structures from satellites?

Research · a month to go · image data, atmospheric sciences, image data



Lyft 3D Object Detection for Autonomous Vehicles

Can you advance the state of the art in 3D object detection?

Featured · a month to go · image data, object detection



RSNA Intracranial Hemorrhage Detection

Identify acute intracranial hemorrhage and its subtypes

Featured · a month to go · health foundations and medical research, image data



Severstal: Steel Defect Detection

Can you detect and classify defects in steel?

Featured · Code Competition · 8 days to go · image data, manufacturing



Kuzushiji Recognition

Opening the door to a thousand years of Japanese culture

Playground · 2 days ago · history, image data, multiclass classification, japan



The 3rd YouTube-8M Video Understanding Challenge

Temporal localization of topics within video

Research · 5 days ago · object detection, video data



Open Images 2019 - Object Detection

Detect objects in varied and complex images

Research · 15 days ago · image processing, image data



Open Images 2019 - Visual Relationship

Detect pairs of objects in particular relationships

Research · 15 days ago · image data, image processing



Open Images 2019 - Instance Segmentation

Outline segmentation masks of objects in images

Research · 15 days ago · image data, object segmentation



Recursion Cellular Image Classification

CellSignal: Disentangling biological signal from experimental noise in cellular ima...

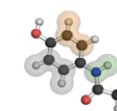
Research · 20 days ago · research, biology, classification, image data



SIIM-ACR Pneumothorax Segmentation

Identify Pneumothorax disease in chest x-rays

Featured · a month ago · image data, object segmentation



Predicting Molecular Properties

Can you measure the magnetic interactions between a pair of atoms?

Featured · 2 months ago · regression, tabular data, chemistry

目录

1/ 数据竞赛介绍 (10min)

2/ 赛题介绍 (15min)

3/ 赛题思路 (15min)

4/ 代码讲解 (15min)

5/ 经验分享 (5min)



更多干货，关注公众号 扫码获取比赛baseline

2 赛题介绍

Competition Introduction

Part2 赛题介绍

Competition Information

赛题名称：NFL Health & Safety - Helmet Assignment

赛题链接：<https://www.kaggle.com/c/nfl-health-and-safety-helmet-assignment/>

赛题任务：在本场NFL头盔追踪比赛中，选手需要构建一个头盔跟踪算法，为测试集视频片段的头盔进行检测。

赛制安排：

- ✓ 2021年8月10日- 开始日期
- ✓ 2021年10月26日- 报名截止日期
- ✓ 2021年10月26日- 团队合并截止日期
- ✓ 2021年11月2日- 最终提交截止日期

Part2 赛题介绍

Competition Information

赛题名称: NFL Health & Safety - Helmet Assignment

赛题任务: 在本场NFL头盔追踪比赛中，选手需要构建一个头盔跟踪算法，为测试集视频片段的头盔进行检测。

赛题数据:



Part2 赛题介绍

Competition Information

赛题名称: NFL Health & Safety - Helmet Assignment

赛题任务: 在本场NFL头盔追踪比赛中，选手需要构建一个头盔跟踪算法，为测试集视频片段的头盔进行检测。

评价方法:

$$WeightedAccuracy = \frac{TotalCorrect_{nonimp} + (TotalCorrect_{imp} * 1000)}{TotalHelmets_{nonimp} + (TotalHelmets_{imp} * 1000)}$$

Where:

- `TotalCorrect_nonimp` is the number of correctly assigned non-definitive impact helmet boxes.
- `TotalCorrect_imp` is the number of correctly assigned definitive impact helmet boxes.
- `TotalHelmets_nonimp` is the total number of non-definitive helmets boxes.
- `TotalHelmets_imp` is the total number of definitive helmet impact boxes.

目录

1/ 数据竞赛介绍 (10min)

2/ 赛题介绍 (15min)

3/ 赛题思路 (15min)

4/ 代码讲解 (15min)

5/ 经验分享 (5min)



更多干货，关注公众号 扫码获取比赛baseline

3 赛题思路

Competition Idea

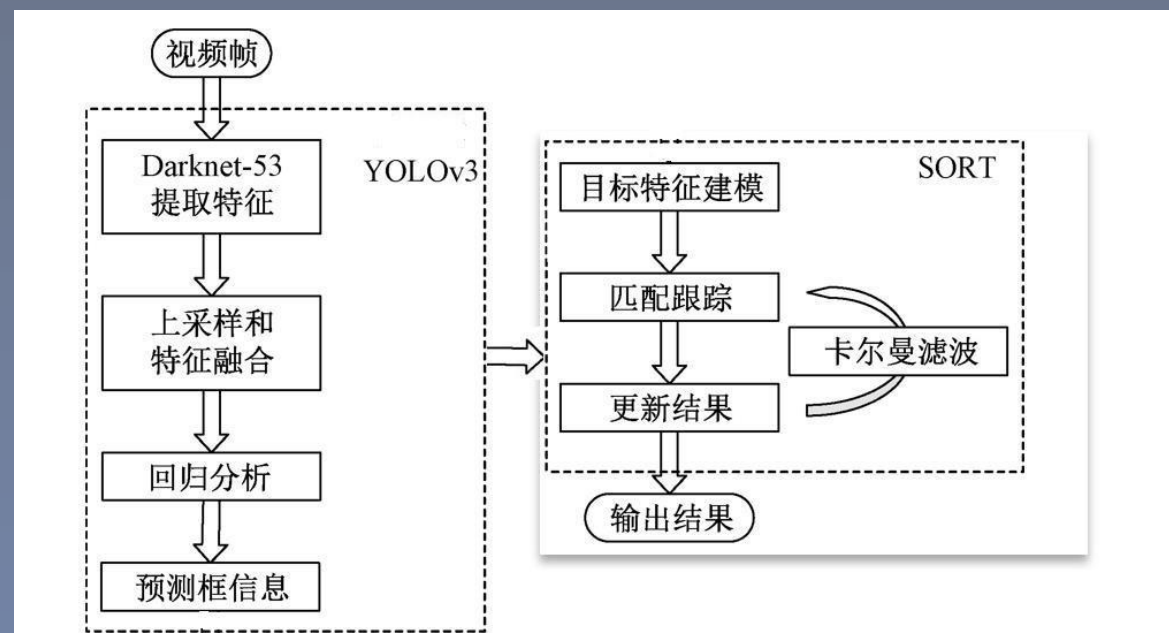
Part3 赛题思路

Competition Idea

赛题任务： 物体跟踪任务

赛题思路： 基于已有的标注数据，训练物体检测 & 目标跟踪模型；

基础模型： DeepSORT + 传感器匹配思路



目录

1/ 数据竞赛介绍 (10min)

2/ 赛题介绍 (15min)

3/ 赛题思路 (15min)

4/ 代码讲解 (15min)

5/ 经验分享 (5min)



更多干货，关注公众号 扫码获取比赛baseline

4 代码讲解

Baseline Code

Part4 代码讲解

Competition Idea

1.代码基础

基础技能包括：Pytorch基础、基础模型搭建模型；

2.代码细节

步骤1：使用DeepSORT进行预测；

步骤2：利用球员传感器数据进行匹配；

目录

1/ 数据竞赛介绍 (10min)

2/ 赛题介绍 (15min)

3/ 赛题思路 (15min)

4/ 代码讲解 (15min)

5/ 经验分享 (5min)



更多干货，关注公众号 扫码获取比赛baseline

5 经验分享

Experience Sharing

Part5 经验分享

Experience Sharing

表格赛	CV赛
任务多样	任务固定
依赖人工	依赖机器和经验
主要是CPU资源	主要是GPU资源

- ✓ 与表格赛相比，CV赛题任务更加固定（分类、分割和检测）；
- ✓ 与表格赛相比，CV赛题需要更多的计算资源，奖金相对多（人少，钱多）；

Part5 经验分享

Experience Sharing

合理使用GPU:

- ✓ Kaggle Kernel平台提供了每周免费的40h GPU时间;
- ✓ Kaggle Kernel平台可以从头训练, 也可以只做预测;
- ✓ 本地1080ti足够参加本地比赛, 做好实验记录;



添加汪汪酱
获取Baseline



Part5 课程安排

Experience Sharing

按时来上课

按时上课，积极互动

每周2次直播，共7次

Kaggle-NFL头盔跟踪				
	课程专题	知识点	作业	时间
1	开营仪式	赛题介绍 Baseline代码详解	跑通比赛baseline, 并完成提交	9/10 20:00
2	多目标跟踪模型流程细节	最终模型细节 模型训练与验证 Baseline进化	使用交叉验证, 对验证集loss进行可视化	9/14 20:00
3	多目标跟踪模型发展	SORT模型 DeepSORT模型 MOTDT模型	掌握追踪算法的搭建和训练	9/17 20:00
4	目标追踪赛题Trick	评价函数 损失函数 训练Trick	掌握目标追踪的数据扩增 方法, 并进行尝试	9/24 20:00
5	计算机视觉赛题复盘	视觉比赛思路总结	掌握视觉赛题的解题步骤, 并尝试给出的目标最终思路。	9/28 20:00
6	模型融合&AutoML	视觉赛题的伪标签 模型集成 预训练思路	掌握目标检测和目标追踪 模型的融合过程。	10/08 20:00
7	比赛冲刺	老师对比赛知识点和上分点 Trick进行整理	老师对比赛知识点和上分点、 Trick进行整理	10/12 20:00

Q&A互动

Ask me anything

Q&A

Ask me anything

■ 赛题不会建模提分？

- ✓ 赛题如何构建验证集？
- ✓ 赛题如何增加模型输入？
- ✓ 文本数据如何处理？

强烈建议参加我们的课程，无痛上手数据竞赛！



Q&A

Ask me anything

接下来做什么？

- ✓ 加入我们的社群。并在平台上完成报名；
- ✓ 获取baseline和运行环境，顺利完成代码提交；
- ✓ 阅读所给的资料，对赛题增加了解；

互动

Fun time

问：你还想学习哪场算法竞赛呢，告诉我们！

前3名回答同学将获得精美礼物~



深度之眼
deepshare.net



你能从本次课程中获得什么

Supplementary teaching



比赛前

赛题指导
+
Baseline思路



比赛中

针对trick的提分思路
不同方案如何进一步优化
针对赛题前沿Paper+资料解析



比赛后

复盘优秀方案+思路
工程&面试应用技巧



你能获得什么

Supplementary teaching

赛题指导+Baseline思路

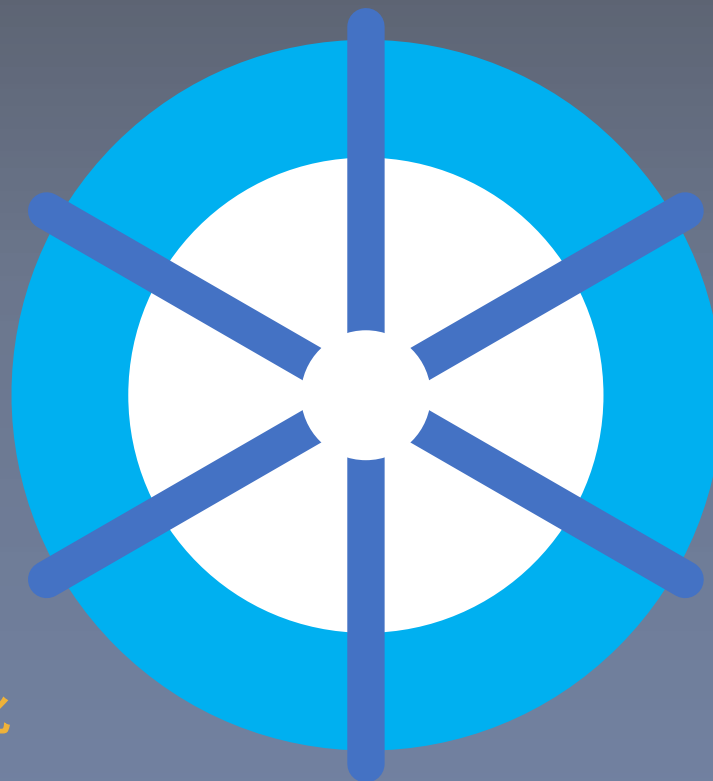
深入解析比赛对应场景，数据，让大家站在出题立场看问题，深入理解老师给出的baseline思路，为什么要这样写，这样写好在哪里，在这个基础上还有什么提升的点。在参加比赛时站在更好的起跑线上。

针对trick的提分思路

根据赛题方向提供通用的提升思路，举一反三，帮助大家不仅在此次比赛中提分，还可以把这些方法迁移到其他比赛中。

不同方案如何进一步优化

先锋队对比赛提前探索，提取Trick。由老师分析每个Trick背后的意义，针对不同Trick，提供下一步优化的方案。



针对赛题前沿Paper+资料解析

拒绝调包侠，通过前沿paper及重要理论的算法原理，不仅让你知道这个方法可以行，更要让你知道背后行的道理。

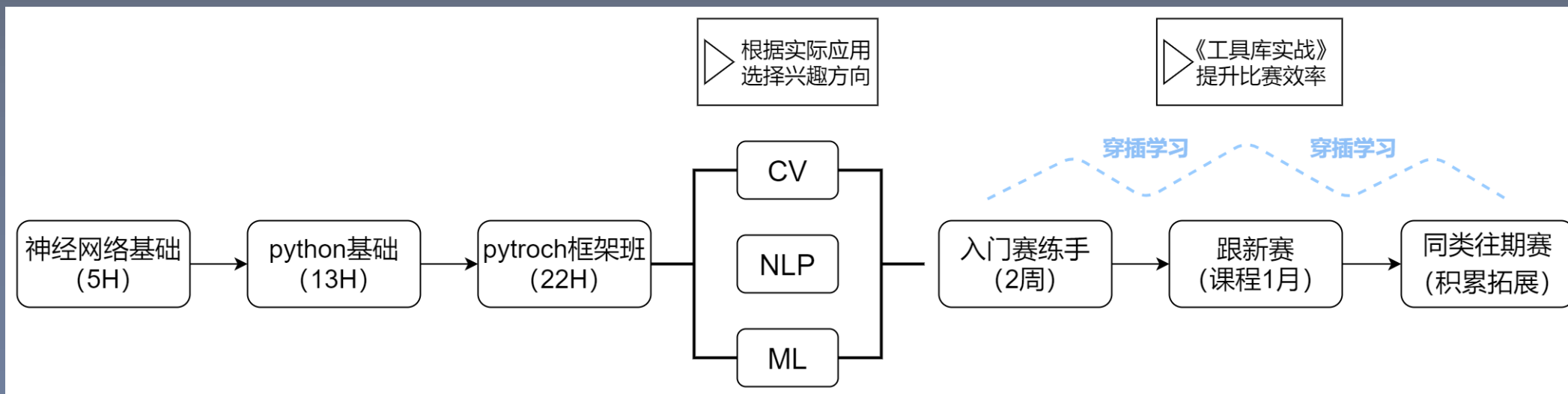
工程&面试应用技巧

除了提分思路分享，老师会分享工业和比赛的异同，比赛思路如何用于工业项目中，比赛经验在面试中的展现。帮助大家在工作或求职中更好地展现自己。

复盘优秀方案+思路

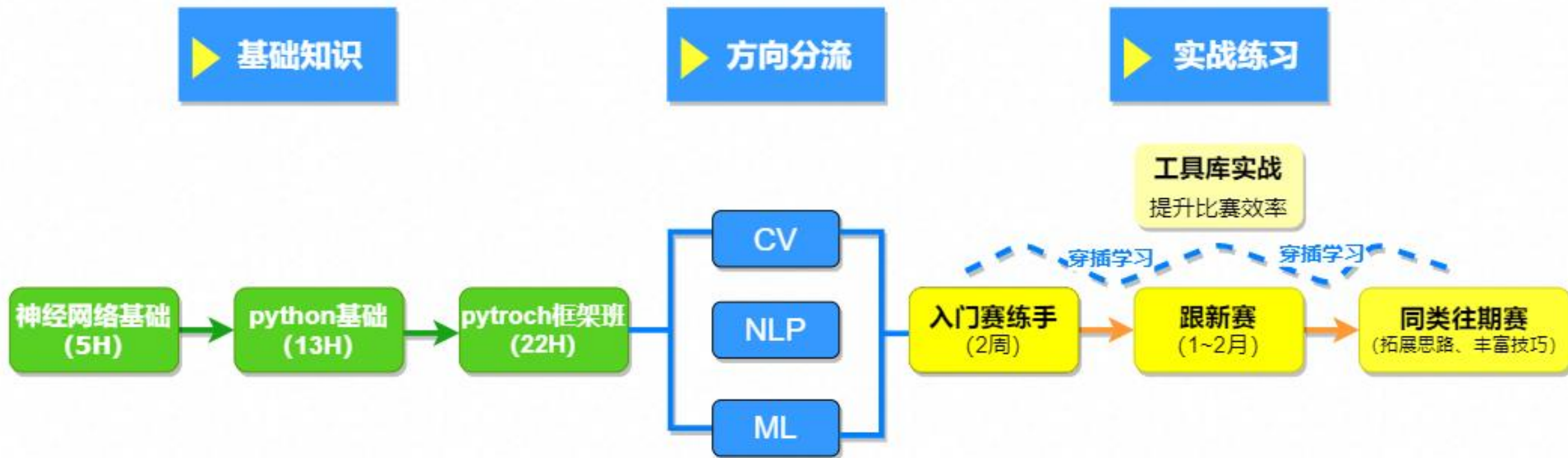
复盘比赛全流程，梳理提分思路的完整逻辑，并对好的开源方案做案例分析。

0 to 大神：比赛会员学习路径



参赛路径

0 to 大神：《比赛会员学习路径》



二十场AI算法经典比赛

机器学习方向

1场比赛

CV方向

6场比赛

NLP方向

6场比赛

时间序列方向

5场比赛

推荐方向

2场比赛

电商多模态

1场比赛

1年带打20±场AI算法大赛

Kaggle



DataFountain

TIANCHI天池

+

含
2000元
基础课

数学基础	298元
神经网络基础	198元
Python基础	198元
pytorch框架班	298元
bert模型班	399元
Python工具库	599元

总价值20000±元



拼团优惠200元

限时1498!

请让我们一起立一个flag!

我承诺:

4周努力上TOP100!



结语

再小的细节，也值得被认真对待

