

第十二届服创大赛 A 类命题企业答疑 5.0

(截止至 3 月 24 日 , 标蓝部分为新增)

【A01】小微企业跨组织人才管理系统【今目标】

1.问题：问题 1

关于评价标准，有什么具体的评价标准吗？会有涉及到什么算法吗？

答：评判标准将主要依据产品设计逻辑的合理性、创意性及运行时的流畅度。不涉及算法部分。

2.问题：项目开始前的数据分析

请问一下，项目的数据是自己收集的吗?(人才对公司的价值和损失)

答：是的，可以自行制作问卷，进行市场可行性分析。

3.问题：开发环境

请问我是在今目标里面开发吗?是只需要在 pc 或者安卓或者 ios 其中一个系统能实现就可以了么

答：不是在今目标内开发。可以在任何环境下开发，pc、安卓、ios 只要有一个系统可以实现即可。

4.问题：今目标是否可以提供接口查询考勤，绩效考核

考勤，绩效考核等是否可以从今目标系统中导入，今目标是否可以提供相关的开发接口

答：本次开发的产品不需要真正和今目标接口对接，但可以预留接口。演示数据可自行杜撰。

5.问题：项目环境

可以用小程序吗

答：可以

6.问题：移动端

请问移动端只实现 Android 端可以吗

答：可以

7.问题：需求问题

该管理系统是只针对员工的信息的后台管理不需要前台的显示信息吗？员工无法查看自己的信息？需要注册和登录吗？

答：皆可，主要取决于产品设计。

8.问题：脱敏处理

脱敏处理针对的是哪些数据应该怎么处理呢？

答：针对某些敏感信息，以实现敏感隐私数据的保护。例如，涉及客户安全数据或一些商业性敏感数据的情况下，就需要对真实数据进行改造或隐藏。比如我们最常见的火车票、电商收货人地址都会对敏感信息做处理

9.问题：关于技术要求与指标问题

请问“在技术要求与指标”中的第五条“可以流畅运行，系统部署于 web 端、pc 客户端、移动端均可；”指的是系统需要在 web 端、pc 端、移动端都可运行，还是三者选其一运行？

答：三选一即可

10.问题：人力资源管理软件

请问这个赛题是今目标有提供给我们软件还是我们自己设计软件，然后规划软件功能呢？

答：需要参赛选手自己设计软件、自己规划软件功能。这也是评分的主要依据。

11.问题：关于接口与开发问题

我想咨询一个问题，不太明白不需要与今目标接口还要跑通是什么意思呢？

因为我的想法是这个赛题希望我们依托今目标实现这样一种功能，在今目标内部存储应聘者信息，然后通过外部开发程序可视化展现，但是没有接口的话，我们自己开发的程序就没办法从今目标调数据吧？

另外，是不是外部开发程序可有可无，主要是用今目标来做，能完成功能即可呢？

如果要自己建一套系统，那我们是否需要自己租用一台服务器以完成演示功能。

答：赛题期望，是开发一个独立的系统，以实现人才管理的诸多需求。理想状态下，是作为今目标功能的扩展，后期植入到产品中，是需要和今目标打通，通过接口导入数据。

但本次比赛中，仅需要有“保留 api 接口”的产品逻辑即可，不需要实际对接接口和数据。

自己搭建的系统，不需要通过服务器完成演示。离线演示即可

1、问题：使用问题

这个除了主张 hr 部门使用，企业的其他部门能否使用

答：可以，但不做强制要求。如果确实设计合理、创意鲜明，可提升印象分。

2、问题：采集信息对象

命题一中被采集信息的人是否可以无工作经验的人

答：作为扩展功能，或许也可以尝试将其用于无工作经验的应届生。扩展为类似校企合作的系统，记录在校表现等。

3、问题：关于与今目标对接问题

老师您好，根据我们团队对贵企业的了解，贵企业已将“绩效管理”，“招聘过程管理”等模块使用在了今目标平台，那么我们后续在做人才管理系统时，如果有与以上模块重复/相似的地方时，我们需要将这些模块自己设计并制作出来还是直接保留接口。如果是保留接口，那么我们在后期制作视频及ppt 时，是否能够提及该模块。

答：保留接口即可。可以提及该模块。这里多提一句，其实今目标内的“绩效管理”、“招聘过程管理”等模块，是依托于今目标核心功能“主线”进行搭建的。可以想象成它是一个灵活、可变的业务流程，并

非固定的系统模块，因此这部分的搭建并无定法。所以，虽然可以近保留接口，但在提及该模块的时候，最好也能阐述一下这部分的搭建思路。

4、问题：关于系统绩效评价内容方面
主管定期对员工进行评价的内容是自己设定吗？有什么具体的内容或要求吗？涉及什么特殊算法吗？

答：可以是主管自己设定，也可以是在模板的基础上进行更改。没有具体要求、也不涉及特殊算法。但是这里可以考虑和今目标系统保留接口，在绩效统计时，直接调用今目标内的数据，例如考勤时间、销售业绩等（仅提供思路，并非最优解）。

5、问题：产品可正常运行跑通即可是什么意思
我们可以借助第三方网站进行制作吗，但这种制作最终显示为：免费域名，可预览访问，功能基本能呈现出来，这样是可以的吗

答：可以。鼓励同学们使用免费资源。该产品并不会直接进行商用

6、问题：有些今目标本身就有的部分要不要做
今目标网站中有些招聘的部分，以及出勤方面已经做得比较完全的话，我们还要重复做一遍吗

答：不需要重复制作，可以保留接口，也可以在 ppt 中提及该模块。详细可参见第 3 个问题的答复。

7、问题：到底题目是中小微企业还是小微企业
到底题目是中小微企业还是小微企业

答：区别不是很大，都可以。事实上今目标产品虽然是面向中小微企业的。但一些上万人的大型企业也会使用今目标，作为其固有系统的补充，或者是供某一项目组使用。

8、问题：A01 题目中“产品跑通即可”是要到什么程度，是我们设计的所有功能都要完全实现吗？

答：能够运行、展示就行，主要也是看产品设计的思路，不需要完全实现

1、问题：现在无法实现的，或者说我觉得实现很难的一些方法，我可以把思路写在文案里面，表现在我的项目中吗？

我想问一下这个背调里面如果涉及到身份信息验证，可以涉及到公安系统吗？如果涉及到我们没有办法链接的部分，我们可以写在文案里吗？因为身份信息是比较私密的，我没验证的话可能会需要涉及到公安系统，我们是没办法去收集这些信息，所以我们可以把它直接设计到里面或者把它放到文案里面吗？

答：可以

2、问题：绩效问题

不同公司的绩效标准如何定？是有标准的嘛

答：绩效标准其实每家公司、每个部门、乃至每个岗位都是不一样的。同一个岗位、不同的管理风格、也会有不同的绩效标准。有些绩效统计就是单纯的结果导向，有些则需要多种要素加权。通常的管理软件，会提供一些可供企业修改的标准化模板。但本次赛题重点并非绩效考核的设计，因此这部分可以进行“黑箱”处理。只需要记录结果就可以，大部分企业的绩效结果都是一个评分或评级。对应的是相应的绩效工资

【A02】基于非特定类别的图像前景主体分割算法【万兴科技】

1.问题：关于最后算法结果的评定

最后关于算法结果评定是光看那 20 多张图的 MIOU 还是会有自己的所有的测试图的 MIOU 来评定，是否需要加入其他指标？因为光是评价主体边缘细节清晰也太主观了。

答：

那 23 张图只是样例图而非测试图，出题方会有专用的测试集进行评估（300+图片）

评价会包含客观评价(MIOU,PA,MSE 等)和主观评价两个部分，包括但不限于上述指标。

2.问题：关于题目要求的半透明物体而测试图里并没有这类图这件事

是不是最后会有关于半透明图的分割测试?

答:

最后会进行半透明分割测试, 具体针对半透明物体的分割要求可详见视频讲解 PPT。

注意: 本算法为通用前景抠图, 适用于任意场景的主体抠图, 不限于 23 张示例图片的主体。

3.问题: 关于网络模型工程化, 包含全部依赖的 C++ 例程

最后的网络模型工程化是指 c++ 用推理框架 (如 openvino、ncnn、tensorrt、tvm) 对训练出来的如 Pytorch 模型进行网络模型部署嘛, 能否稍微介绍一下方法。

答:

是的, 用上述 OpenVINO/MNN/TensorRT/TVM 等任何一种推理框架部署, 具体可以去 OpenVINO 等推理框架的 github/官网上去搜。

本题中要求统一用 C++ 进行部署, 方便横向对比性能。

4.问题: 测试图片大小差别大不大

最后需要进行测试的图片大小差距大不大, 是不是就是按照常规情况进行设计。

答:

测试图片分辨率不做限制, 任意分辨率。

5.问题: 关于训练集是小样本还是大量数据

出题方最后想要得到的是一个基于小样本而训练出来的符合这 20 多张图情况的模型, 还是基于大量数据的, general 的一个模型

答:

需要一个比较 general 的模型, 也就是泛化性能好鲁棒性强的模型。

6.问题: 图片压缩

是否允许对给的测试集图片先进行压缩, 之后再行分割, 这样时间会更短

答:

可自行选择是否压缩, 但最终输出结果分辨率应与原始分辨率相同。

图像可以压缩 inference，但是最后的边缘锯齿等问题需要解决，因为最后的评定标准包括了边缘的抠图效果。

7.问题：可以使用 python 嘛，需不需要进行前端展示
可以使用 python 嘛，需不需要进行前端展示

答：

可以使用 python，开发语言不限，但是部署语言要求 C/C++，并且要能进行前端展示。

8.问题：模型大小的限制问题

模型大小不超过 100M，指的是参数量小于 100M，还是 pth 文件大小？

答：100M 指的最终提供的模型大小

9.问题：关于运行系统

运行的系统是 Windows 还是 Linux

答：Windows

10.问题：关于输出结果

黑白的 MASK 是否需要输出

答：需要

11.问题：关于“算法性能指标：在 1080 TI GPU 上处理一张图片的时间不超过 5 秒”提问

最终评判性能的指标，是指整个程序运行的时间(加载模型、图片预处理、推理、图片后处理的总时间)不超过 5 秒，还是指模型推理的时间(一张图片)不超过 5 秒？

另外，最后需要提交的可执行文件，是整个工程代码(例如整个 vs 工程)，还是可执行的 exe 即可？
因为会涉及到不同的配置环境，仅提交 exe 不能保证可以运行。

答：整个程序运行时间；最后需要提交独立可执行的应用程序。

12.问题：关于开源代码和交互界面

做一个 pc 端可执行程序，里面核心算法可以使用开源代码吗，使用的话会标注出来，然后在外围加上我们自己做的交互界面

答：可以

13.问题：A 组赛题答疑 V1.0 中的 A02 的问题 7 提到的的前端是指 PC 的 GUI 界面还是 Web 前端
由于赛题手册指出需要提交 PC 离线可执行文件并且不能联网，但是上述问题 7 的企业解答却提到需要“并且要能进行前端展示。”，请问这个前端是传统的 Web 前端还是特指 PC 端的 GUI 界面呢？

答：GUI

14.问题：前端设计

请问前端设计需要设计的很完善吗？还是只给出一个简单的交互界面就可以了，如果是简单的交互界面会减少最后的评分吗？

答：不会，只给出简单的交互界面，保证相关功能的正常使用即可。交互界面的美观和易用程度不会作为评判标准，但如果因交互问题导致功能无法使用，会扣分甚至成绩无效。

15.问题：深度学习

您好，我想问一下，如果是使用语义分割算法的话，模型大小限制在 100M，感觉就是要求比较严苛，用传统的 resnet 和 darknet 都基本过百，只有 mobilenet 的比较合适，但效果就没有这么好。

想确认一下我们团队是否走进了误区呢？

答：模型小于 100 是完全可以做到的，比如对模型进行优化压缩等

16.问题：框架语言

可以使用 python 的神经网络模型吗

答：语言没有限制，最终提交符合要求的可执行程序即可

17.问题：使用显著性检测模型实现分割还是用深度学习模型实现分割，还是用深度学习实现显著性检测后再来分割？

使用显著性检测模型实现分割还是用深度学习模型实现分割，还是用深度学习实现显著性检测后再来分割？

答：不会具体规定如何实现，重点看对各种图片类型分割的效果，最终提交符合要求的可执行程序即可

18.问题：框架搭建

可以使用 opencv 吗？还是必须用 python+tensorflow 来自己搭建框架？谢谢

答：可以，不做限制

19.问题：算法

1.算法是否要求自己写或者是否可以使用开源的

2.框架是否可以使用 Keras

答：1.都可以 2.可以使用 keras

1、问题：提交程序版本

请问提交程序可以交 Ubuntu 版本的嘛？

答：最终打包成.exe 可执行程序即可

2、问题：对于作品要求的问题

1.模型是必须自己设计吗（不能使用已经已经发表的论文中的 SOTA 方法）？

答：均可，不做强制要求

2.模型的评判标准是怎么样？只要求精准度度嘛还是会从速度与精准度两方面综合考量？如果想设计一个轻量模型是否会吃亏

答：会从速度与精准度两方面综合考量

3、问题：有关 A02 抠图对象的选取问题——单体与个体前

对于讲解视频中提到的,主体背景可以是个体,也可以是多个个体,是否意味,图片中心区域较显著的目标是一条狗坐在一个家庭之前,而赛题要求是只抠出有关狗的图像即可?

答：狗和人都要抠



4、问题：关于输出图片的分辨率

老师您好，赛方在答疑 2.0 中提到，输出结果的分辨率应与原始分辨率相同，那如果是 gui 展示里分辨率也是原始分辨率的话，可能会导致整个界面过于 flexible，影响美观，是否可以固定 gui 大小，在 gui 中展示 resize 后的图片，原始分辨率的图片通过另存为其他目录的方式呈现？

答：gui 展示可以适当缩放图片大小，但是 mask 结果必须和原始分辨率相同

1、问题：关于模型部署的问题

请问可以用 pyqt5 和 pyinstaller 打包成 exe 文件吗？

答：可以，最后也要提供依赖的资源文件，要独立可运行。

2、问题：模型大小问题

赛题手册中指出模型不超过 100M，能否使用多个模型适应不同的识别对象并且每个单独的模型不超过 100M，还有这个 100M 是指 C++部署后的文件大小吗？

答：不能使用多个模型适应不同的识别对象，100M 是指 C++部署后的模型大小

【A03】数据资产估值定价分析【浪潮卓数】

1.问题：关于算法的要求

我注意到我们算法要求使用 RNN，意思是只能使用循环神经网络吗？不可以使用其他的手段包括机器学习（lightgbm, xgboost）或者其他的深度学习手段（lstm），因为京东万象的数据集十分有限，差不多一千左右，其实用深度学习方法效果达到 90%基本上不具备什么泛化能力

答：

可以使用其他算法，我们也希望学生用多样的算法来实现，当时审题老师觉得不限定算法比较难，我们才加了限定

2.问题：关于展示的选择

请问咱们展示只能是 web 端嘛？是否可以移动端？

答：

展示也可以用移动端。

3.问题：系统输出的价值

比如利用市场法，最终输出的价值是最后的标价吗？

是交易的建议价格还是数据资产的价值货币化

答：基于市场法实现的两个是一回事。

4.问题：模型和方法

采用的深度学习模型，必须是循环神经网络吗，还有必须要用市场法做吗

答：当前基于市场法的循环神经网络模型相对比较成熟，同学们实现起来较为容易，如果同学有能力基于其他算法模型实现也可以。

5.问题：api

api 的接口算不算数据

答：算。

6.问题：数据抓取器

这个数据抓取器最后会嵌入到我们所做的作品里吗？

答：重点考察算法模型，数据采集是有则优，其实做同学们如果自己做真实案例项目，数据一般是自己抓取的，建议大家都做一下数据采集这块。

7.问题：关于数据集和数据结果的要求

我们需要对京东万象上哪些数据进行数据资产评估，因为感觉京东万象上很多数据并不能进行数据资产评估。同时在进行数据资产评估时，不可能仅仅通过一个算法解决所有类型的数据资产评估，因为数据资产包括很多类型的数据（电商数据、风投数据等），我们具体需要对哪些类型的数据进行资产评估？同时评估的结果需要进行验证，验证的结果从哪里获取？因为我看到京东万象上只有 API 每次的价格

答：不是只有 API 数据，还有块数据，数据类型可以从企业数据入手。

8.问题：关于数据的问题

题目会提供数据吗，比如测试数据和训练数据，看题目有准确率上的要求，应该是有测试集的吧

答：不提供训练集合测试集，需要同学自己从互联网获取，建议是从京东万象获取（数据较全面、有真实成交量），如果使用爬虫，请不要影响网站正常访问。

1.问题：网站交互

网站的交互可不可以复杂一些，可否在页面设置关于数据的参数值（如预期收益等）的输入，而不是只输入一个数据包

答：网站设计可以复杂，输入数据可以包含数据资产名称、数据维度（电商、企业等）、数据条数、字段数量等。

2.问题：数据包内容

考核输入的数据包中一般包含哪些内容，有哪些特征值。是关于数据交易的特征值（如交易量、浏览量）还是关于数据本身的特征值（如成本）

答：鉴于同学们较难接触到数据本身的特征值（如数据的质量、成本等特征），我们建议同学们以市场法入手进行分析，即关注数据交易的特征值，输入数据可以包含数据资产名称、数据维度（电商、企业等）、数据条数、字段数量等。

3.问题：数据接入

带评估的数据是以什么形式接入的，需要手动输入吗

答：待评估的数据输入的是数据的特征值，而不是数据本身。

4.问题：关于“准确率”的问题

我们注意到这是一个数据资产评估的问题，那么我们最后应该是基于数据特征对数据价值做预测，那么应该是一个回归问题而非分类问题，请问题目中说明的“90%准确率”是什么意思？对于回归问题的 accuracy score 应该如何定义呢？

答：感谢这位同学提出的问题，我们对技术指标准确率作出调整，使用通用评估指标均方根对数误差 RMSLE 来进行评估。

5.问题：数据类型

请问数据类型大致是什么样的

答：待评估的数据输入的是数据的特征值，而不是数据本身。

6.问题：Web

请问设计 web 需要对外发布吗

答：如果对外发布是公网访问的话，则不需要，但是至少需要自己的电脑访问演示。

1、问题：评分标准

我想请问一下我们做的模型是关于回归问题的一个模型，那么我们的评分标准是什么呢？你们文档上面说分数要到达 90 以上。这个分数是值什么评分标准算出来的数值？R 的平方吗？还是其他的，或者如果用 mae 来作为模型的评价指标，那么这个 mae 怎么比较呢？因为大家的数据集都不一样，数据的方差、平均值也就不一样。

答：同学你好，评分标准已经变更，使用通用评估指标均方根对数误差 RMSLE 来进行评估。

2、问题：数据来源

京东万象的数据平台上数据较差，使用爬虫方法很多网站都反爬，数据难以获得

答：使用市场法的情况下京东万象已经是现在相对成熟的数据交易市场，建议以市场法进行分析，如果有其他数据来源渠道，不排斥使用其他模型算法。

3、问题：能否创新一种交易模式？用于解决一些目前数据资产交易的痛点问题以便给予这项资产更合理更灵活的估值。

比如创造出一种交易系统能够解决数据交易中数据可能会未经所有方允许的被交易给第三方平台，这样数据资产的一些属性（如交易次数较多使得数据资产质量下降）对估值的影响会更加小。这样的方法会不会被认定为离题呢？

答：从本次赛题来看是离题的，从课题研究方面来看还是鼓励大家有新的想法，目前在数据交易上还有很多问题需要明确和解决，现在没法确定谁的模式就是正确的。

【A04】航空旅客付费选座意愿识别【东软】

1.问题：竞赛用数据集中 emd_label 的含义

数据集中出现了 emd_label 和 emd_label2，其中后者是目标列，前者给的说明是 emd,我们团队对 emd 的含义存在疑惑。

答：

业务上有两种规则用来确定是否为付费选座行为。用 emd_label1 确认的付费选座范围较大，emd_label2 被包含在其中。

本次竞赛中采用 emd_label2 为付费选座业务准则。

2.问题：“用这些特征描述付费选座旅客的画像”是指眼睛鼻子那些的还是抽象的行为模型？（5.项目说明中的【问题说明】中，提到，“用这些特征描述付费选座旅客的画像”。请问这个画像是指眼睛鼻子那些的还是抽象的行为模型？）

答：

“付费选座旅客的画像”在这里特指航空旅客的业务行为特点。用旅客已经发生的业务行为的期间内统计结果，描述旅客在航空业务中的共性和特性。

3.问题：数据集中某特征的用处

数据集中字段名为 emd_lable2 的特征为目标列，但还有一个字段名为 emd_lable 的一列数据，请问这一列数据是什么用呢？

答：

同 1.

4.问题：提供的数据有问题

按照赛题视频的介绍，数据项里面 6 个月的数据是包括 3 个月的数据，但是为什么有的列 y1 即第一年有数据，y2 即第二年数据为 0，y3 即第三年有数据呢？如果按照说明，第二年是包括第一年的数据的，不应该出现这样的情况。这样的列包括 JL 到 JN 列、JO 到 JQ 列、RE 到 RI 列。

答：

你的指摘是正确的，这部分 y2 的统计结果为空，系样本数据预处理时的错误。

5.问题：航班日期

数据特征列的 航班日期 指的是截止数据统计结束时的最后一次坐飞机时间还是什么意思？

与最近一次出行日期这一特征列有什么区别？

答：

数据列”航班日期“指，旅客搭乘的航班的起飞时间。如果同一个旅客搭乘多次飞机，会有多条乘机记录和相应的航班日期。

“最近一次出行日期”指，同一旅客在统计期间内的最大的航班日期，Max(航班日期，统计期间)。此列数据部分为空，系样本数据预处理时的错误。

6.问题：pref_aircraft 那些列 73h

pref_aircraft_那些列 (好几列) 出现很多 73H 这个数据表示什么，是导出错误吗

答：字段 Pref_aircraft 是旅客搭乘次数较多的机型。如 73H 指波音 737.

7.问题：题目问题

测试集数据 7000 条和验证集数据 6000 多条是一个东西么，还是会后期再给？验证集 emd_lable2 全部为 0，是没有意愿还是意愿未知？

答：测试集数据 7000 条和验证集数据 6000 多条是同一个东西。

emd_lable2 字段是已知付费旅客标识。这个字段在验证集中全部为空值 (0)，需要通过模型预测得出。

8.问题：关于意愿识别的设计开发

请问关于航空旅客付费选座意愿识别需要开发输入数据输出结果的图形化界面吗？还是只需要完成对测试数据的结果输出，一个模型就可以了？

答：不需要提交图形化界面。只需提供预测的结果和模型解释即可。

9.问题：数据接口

请问可以提供连接 Java 和 excel 的数据接口吗

答：不可以。

10.问题：指标含义不明确

1.请官方解释一下“距离下次飞行的天数”，这个指标是怎么产生的，下次飞行是哪一次飞行

答：“距离下次飞行的天数”指：在统计的当时时点，如果旅客有未出行的机票，则计算公式为：“距离下次飞行的天数=未出行机票的起飞时间 - 统计时点”。

11.请官方解释一下“最近一次出行日期”与 seg_dep_time 的关系

答：“seg_dep_time”为某航班的出发日期。“最近一次出行日期”为在统计的当时时点，这个旅客的上一次乘机日期。

12.问题：指标数据不明确

请官方解释一下“最近一次出行日期”中数据为 0 是什么含义，是数据缺失吗

答：这是数据整合时的问题。以提供的数据为准，请酌情使用。

13.问题：指标数据不明确

请官方解释一下“距离下次飞行天数”为负数是什么含义

答：“距离下次飞行天数”= 0，指统计时点时，该旅客没有未来的出行计划；

“距离下次飞行天数”> 0，指统计时点时，该旅客距离下次飞行天数；

“距离下次飞行天数”= -1，指除以上两种情况以外的业务状态。

14.问题：数据之间相互冲突

航班记录 seg_dep_time 显示 2017，最近一次出行 recent_flight_day 显示 2018，最近 Y 年飞行次数 flight_cnt_y3 显示 0，这种情况的数据有 5978 个，为什么都有飞行记录，然后最近三年的飞行次数却为 0

答：这是数据整合时的问题。以提供的数据为准，请酌情使用。

15.问题：指标含义不明确

请问官方“最近”是从哪一天看是算的，是从航班记录 seg_dep_time 开始往前推三年的吗

答：不是。“最近”是一个统计快照的概念，是从统计的当时时点开始往前推。所以，“最近三年”这个指标你可以认为是“Today () - 3 年”。

16.问题：指标含义不明确

之前答疑的时候，说“最近”和“过去”是同一个含义。但是我们队伍对比两个数据，发现两者的数据根本不统一，所以我们团队对这个指标还是存在疑问，请官方解释

答：“最近”和“过去”在统计时间段上一致的，都是从现在开始往前推若干时长。
如果你指的是字段“最近 Y 年飞行次数”和“过去 Y 年总飞行次数”，这两个指标有不同的业务含义，所以数据不一致，这两个名字确实有歧义。

1.问题：指标理解不清

官方提供了某一次的航班信息，以及后面一系列的指标。请问后面的指标的统计结果是在该次航班之前的信息，还是该次航班之后的信息

答：要进行具体的分析，看给出的信息的具体含义，来判断是之前航班信息还是之后的航班信息。

2.问题：航线数目

题目中有提到过总共十六条航线，但在数据里常飞航线远不止十六条，这里的十六条航线是怎么划分的呢

答：训练数据中航线只有 AB1001-AB1016

3.问题：影响力标准

mdl_influence影响力 在表格中是有数据的，能提供一下评分标准吗

答：主要根据客户购买的舱位的等级、购买的次数以及所选服务等。

4.问题：特征因子概念

tkl_avg_interval最近 Y 年平均订票时间间隔 能说明一下大部分单位是什么，数据如何处理的

答：m3 m6 y1 y2 y3，指累计三个月，半年、一年、两年、三年的间隔，以小时为单位。

5.问题：旅客价值

mdl_mcv旅客价值 如何判断旅客价值的，如何得出的评分

答：旅客价值根据其购票数量以及购票同时选择选座服务的次数决定。

6.问题：关于数据的缺失情况

数据里的那些 0 是由于保密原因而要隐藏的，是只有这种类型的‘0’，还是在次数、里程等一些数值型数据里也有隐藏

答：这是数据整合时的问题。以提供的数据为准，请酌情使用。

7.问题：有关 pref_orig (偏好出发地) 数据的问题

请问例如：pref_orig_y1_1、pref_orig_y1_2 中 y1 后面的数字后缀代表的是什么意思？

答：偏好出发地根据次数来进行了排序。

8.问题：flt_delay_time 中出现了大量负数据

在 flt_delay_time 中出现的大量负的数据，适当错误数据处理吗？

答：数据整合时的问题，可以当错误数据处理。

9.问题：特征业务含义

字段“最近 Y 年飞行次数”和“过去 Y 年总飞行次数”，这两个指标有不同的业务含义，所以数据不一致，这两个名字确实有歧义，请问具体的业务含义分别是什么呢？这将有助于我们的特征衍生工作

答：“最近”和“过去”在统计时间段上一致的，都是从现在开始往前推若干时长。如果你指的是字段“最近 Y 年飞行次数”和“过去 Y 年总飞行次数”，这两个指标有不同的业务含义，所以数据不一致，这两个名字确实有歧义。

10.问题：机票平均消费的特征

tkt_avg_amt_y3 最近 3 年机票平均消费与 tkt_avg_amt 机票平均消费是否意义一致？

答：数据整合出现问题，以 tkt_avg_amt_y3 为准。

11.问题：特征含义问题

pit_next_level_dist距下一会员级别的会员积分差

pit_next_level_leg距下一会员级别的行段数

这两个字段存在 0 值与负数（如-100，-99），请问负数时有什么业务含义

答：数据整合时的问题，可以当错误数据处理。

12.问题：城市与省份数据是航班的出发地还是目的地？

city_name province_name nation_name 这三个特征是本次航班的特征还是用户个人所在地的特征，这将有助于我们的特征衍生工作

答：是用户个人的特征，不是航班特征。

13.问题：预测结果旅客名单需要去除重复旅客编号吗？
赛题要求提供预测结果旅客名单，名单中需要去除重复旅客吗？

答：需要去除。

14.问题：属性名代表含义
请问属性中的***_y1_1,***_y1_2, 后面的 1 和 2 是什么意思？例如：
pref_month_y1_1 和 pref_month_y1_2, pref_month_y1_3 分别代表什么意思，
pref_month_y1 代表累计一年常飞月分

答：对次数进行统计并排序后得到的特征。

15.问题：member_level 特征内容意思不太明确
这个特征是会员等级，但经过百度，那些表示等级的英文缩写查不到，所以不知道那些到底哪个是高等级，那个是低等级

答：会员分出等级，具体不予解释

1、问题：验证集标签全为 0，模型好坏如何评判

您好，我想问下模型最后的评判标准是模型在验证集上的准确率和 auc，还是在训练集的准确率和 auc。因为从验证集上来看其标签阳性比率为 0，同时训练集标签阴阳比率不平衡。如果不进行处理，模型最终对于 0 标签的权重会非常大，这样会导致训练好的模型在验证集上的准确率会高达百分百。所以我想问模型最后评判是基于什么条件，验证集是否需要调整

答：提交要求：提供不超过 500 人的预测结果旅客名单，按照有付费选座意愿的概率倒排序。验证集的标签是填充的，按照没有值来处理，是需要通过模型来计算的。现实中数据不平衡的问题很常见，在模型训练时要注意构建训练集的比例。

2、问题：数据之间相互冲突

最近 3 年飞行总里程数为 0 的数据有 5588 个，但是其中最近 3 年飞行次数为 0 的数据只有 871 个，想问一下既然存在飞行次数，那么为什么飞行的总里程数却没有

答：这是数据整合时出现的问题。以提供的数据为准，请酌情使用

3、问题：对指标含义存在疑问

指标“pref_aircraft_m3_1”中最后一个数字“1”代表什么

答：对偏好进行统计后，会按照喜好程度高低进行排序，依次递减。

4、问题：对会员级别的疑问

会员级别中的 STD,CHD,SILHL 等有什么含义

答：是会员的编号，为了将会员进行等级划分，编码成数字后直接使用即可，具体不予解释。

1、问题：emd_lable 的问题

emd_lable 字段在训练的时候能否保留作为一个特征字段

答：不可以。

2、问题：样本质量描述

在原数据里，出现了相同的“顾客姓名”、“护照”、“航班日期”，但“机票费”不一样的样本，是不是理解成顾客买了两张同样时间出发的飞机票，这两飞机票价钱不一样

答：这是数据整合时出现的问题，以提供数据为准，请酌情使用。

3、问题：特征描述问题

在原数据里，有两项特征为“机票费”，“机票税费”，想问下“机票费”是否包含“机票税费”，按照网上的定义，“机票费”已经包括了“机票税费”，但是在原数据里出现了“机票费”接近 0，而“机票税费”高达 200+的样本

答：这是数据整合时出现的问题，如果与常识不符可以认为是质量差的数据，不参与训练。

4、问题：特征描述问题

在那些选了加钱选座位（emd_label2=1）的样本中，“机票费”是否包含“加钱选座位”的费用

答：机票费中不包含选座另外花费的费用，另外花费的费用来选座位有一个单独字段。

5、问题：样本描述

为什么有些样本，“会员编号”不为 0，但“会员等级”却为 0；同时，为什么有些样本，“会员等级”不为 0，“会员编号”却为 0；

我们感到疑惑的是，若该样本的“会员等级”\“会员等级”不为 0，则意味着该样本的身份是会员，其“会员等级”\“会员等级”不应该为 0；

答：会员编号与会员等级从 0 开始编码，区分出会员的不同类型即可。

6、问题：旅客价值中的数据是什么意思？越大表示价值越大吗？

10、20、30。。。250、300 这些数据是越大表示旅客价值越大吗？

答：旅客价值是按照选座业务逻辑进行的编码，为了区分价值的不同，与数字本身意义无关，请酌情使用。

7、问题：属性名代表含义

属性后面的***_y3_d1,***_y3_d30 是什么意思

答：对偏好等进行统计并降序排序。

8、问题：关于模型准确度的疑问

“识别出的有付费选座意愿的旅客，准确率越高越好”这句话的准确率是指（1）正确预测有无付费选座意愿的数量/需要预测人员的总数 还是指（2）预测有选座意愿的人的数量/具有选座意愿的人的总数 还是（3）其它？

答：根据本数据的正负样本情况，可以查看模型的 auc 指标来进行判断模型的好坏。

【A05】智能网联汽车辅助驾驶安全信息检测系统【东软】

1.问题：车型信息和行人信息识别

识别车型信息，是指识别车辆的头部、后部、相对距离等信息嘛；

识别行人信息，是指形成行人识别框图，标明行人位置以提醒驾驶员嘛？

答：识别车型信息：是指非机动车，机动车等，车型信息具体更好。

识别行人信息，是指行人识别框图，提醒驾驶员。

2.问题：应用设备

本项目所需应用的设备是中控和移动设备均需实现；还是二选一实现

答：二者选一即可

1.问题：关于手机端所需显示的检测信息详情

请问车头识别到并传回的障碍物信息包括障碍物大致类型、障碍物图像吗？

车尾识别并传回的障碍物信息包括障碍物大致类型、障碍物图像、盲区实时视频画面吗？

答：1、包括障碍物大致类型、障碍物图像。

2、车尾识别需要传回障碍物信息包括障碍物大致类型、障碍物图像、盲区实时视频画面。

2.问题：系统完整性

请问本赛题是否需要实现一个完整的应用系统，还是就实现检测识别并进行反馈即可？

答：不需要完整系统，实现检测识别并进行反馈。

3.问题：登录与商化

请问应用系统实现的功能需要进行用户登录，这么做会影响商业化嘛

答：登录不是必须的，需要实现检测识别并进行反馈。

4.问题：接口调用

其他一些功能的实现能否调用第三方接口？

答：可以调用，备注好出处即可。

1、问题：实时性

检测需要实时对摄像头传来的画面进行分析吗，还是可以对已有的视频进行检测？然后如果要求实时性是否可以接受一定程度的延迟？比如 2-3s？

答：检测需要实时对摄像头传来的画面进行分析可以接受一定程度的延迟，请酌情优化，2-3秒有点迟钝。

2、问题：数据传输距离

数据接收和显示的终端离设备的距离是否有要求，如果有是多少？

答：没有要求

3、问题：连接方式

终端可以通过哪些方式连接（摄像头）设备，是否有限制？

答：方式自定，不限制

4、问题：显示设备

可以使用 QT 在边缘计算平台上开发界面，当作显示平台吗？

答：可以，平台不限制

5、问题：障碍物类型

车尾盲区检测所需要识别的障碍物是指车辆、行人等障碍物还是指石头、栏杆等障碍物？

答：这些统一都称为障碍物。

【A06】基于 AI 学习白流量过滤【深信服】

1.问题：关于数据集的问题

请问本题(白流量过滤)是否会提供数据？

答：

企业不提供数据，学生可以在自己的网络出口抓取数据包

2.关于数据集的问题，官方说学生可以在自己的网络出口抓取数据包，那请问能否直接使用一些已有的网络数据集（如 CTU-13,Malware-Traffic-Analysis.net）

答：本课题的主旨是对网络中的白流量进行过滤，对过滤后的流量再进行分析，不能直接使用恶意流量数据集。重点在白流量的识别和过滤，在不使用恶意软件或遭受攻击的情况下，学生自己的网络出口流量基本全为白流量，这些流量可以作为实验的数据。

3.在任务清单的第三点中，我们要提供解包分析和白流量过滤

其中，关于解包分析和白流量过滤有不太理解的地方，希望官方可以做出解释：

（1）解包分析：这里的解包分析是说，对于过滤完白流量之后的可疑流量，我们还需要像全流量检测一样编写解包、还原等部分的代码，还是说这部分的代码企业是会提供给我们的，亦或者是解包分析不是这个意思？

答：本课题的目标是过滤白流量，过滤后的流量可能会存在恶意流量，需要对过滤后的流量进行解包和分析。对过滤后流量的分析不是本课题的重点，同学们可以将重点放在对白流量的识别和过滤上。

（2）白流量过滤：这里的白流量过滤我的理解就是本命题的主要工作，将白流量过滤出来，然后看过滤过程的准确率，耗费时间等，不知道我的理解是否正确？

答：是的，评估的标准主要是白流量过滤的准确率，消耗的时间、CPU 和内存等。

4.请问官方如何评估我们算法的效果，用什么评估指标？是不是用已标注好的测试集来进行评估？

答：评估的标准主要是白流量过滤的准确率，消耗的时间、CPU 和内存等。测试集需要用标记好的数据，但不限于具体的流量范围。比如使用 A 网口的数据进行训练，但会使用 B 网口的流量进行测试。

5.命题上说我们要学习白流量进行过滤，请问我们能否换一种思路，从识别可疑流量的角度来进行，与可疑流量相对的就是异常流量？还是说应该直接给出白流量的具体特征？

答：因为从全流量中识别可疑或者恶意流量效率较低，所以希望能先将白流量过滤掉，再对剩下的可疑流量进行分析。如果直接从可疑流量的角度进行，会和本课题的初衷相悖。

6.赛题上说的【开发工具与数据接口】是 Python、Anaconda、TensorFlow，请问这个是指定好不能再修改了吗？

答：可以修改，同学们选择熟悉的编程语言和平台即可。

7.问题：关于加上流量过滤功能后，性能较全流量检测提升一倍以上，如何衡量
请问下若项目仅实现流量过滤功能，但因为官方没有提供全流量检测 API，不知如何衡量流量过滤模型给全流量检测性能的优化效果。

答：本课题的主旨是对网络中的白流量进行过滤，对过滤后的流量再进行分析。流量过滤之前安全设备需要对所有的流量都进行检测，但白流量占据了所有流量的绝大部分，因此，希望将白流量先过滤掉，然后再检测剩下的可疑流量。性能提升一倍以上，也是指流量过滤之前和过滤之后对可疑流量检测的对比。重点是对白流量的过滤。关于恶意流量，建议使用一些常见的木马样本来产生数据。课题效果评估在真实的网络环境中测试即可。

8.问题：前端部分

对于用来体现功能地网页前端内容及功能有要求或者限制吗？

答：本课题仅限于流量过滤功能，不需要做前端设计。

9.问题：信息中提到的安全设备解包分析，那么这个环境具体是指哪种环境？

需要解包分析的安全设备是什么设备，所处环境是怎样

答：环境可理解为常见的网络环境，及实验室或学校网口数据。解包分析设备为网络边界检测设备，例如防火墙。

10.问题：请问该赛题最终成果需要以什么形式呈现？

请问最终该赛题成果需要部署在哪些地方？以何种方式部署？也就是说我们提供模型算法后，需要以何种形式来使算法得到应用？

答：成果已方法论和验证原型的方式呈现。

11.问题：比赛范围

是同时处理 HTTP 和 DNS 还是选择其中一种处理，还是可以考虑其他的流量包也可以

答：白流量范围不仅包括 HTTP 和 DNS 流量，也包括常见的音视频和交互软件流量，即真实环境中的流量。

12.问题：白流量

对于白流量需要进行二次分类吗？

答：不需要。

13.问题：关于算法性能提升的疑问

赛题技术要求与指标的第三点说：加上白流量过滤功能之后，整体性能较全流量检测提升 1 倍以上，包括检测时间、检测占用 CPU 和内存。

这里有一个疑问，怎么判断整体性能较全流量检测提升 1 倍以上呢？全流量检测的性能是如何评估的呢？

答：本课题的主旨是对网络中的白流量进行过滤，对过滤后的流量再进行分析。流量过滤之前安全设备需要对所有的流量都进行检测，但白流量占据了所有流量的绝大部分，因此，希望将白流量先过滤掉，然后再检测剩下的可疑流量。性能提升一倍以上，也是指流量过滤之前和过滤之后对可疑流量检测的对比。重点是对白流量的过滤。关于恶意流量，建议使用一些常见的木马样本来产生数据。课题效果评估在真实的网络环境中测试即可。

14.问题：流量包解包

您好，我想大致了解一下，深信服产品的全流量检测的大致流程，以及解包的一些大致技术，不知这些是否方便介绍。麻烦了

答：抱歉，涉及公司技术路线，不能向外提供。

15.问题：数据示例

您好，出题方能否方便提供一些检测数据示例，未进行解包的示例能否方便提供。在这个方面有一些疑问

答：课题重点在于对白流量的过滤，采用实验室网口真实数据即可。关于恶意流量，建议使用一些常见的木马样本来产生数据。

16.问题：具体恶意流量比例

请问贵公司在进行白流量过滤时，一般采取的样本中的 DDos，恶意爬虫，木马等恶意流量分布比例是多少？我们团队想以此为依据，选取适合的数据集，设计出更加符合公司业务要求的算法模型。

答：课题重点在于对白流量的过滤，采用实验室网口真实数据即可。关于恶意流量，建议使用一些常见的木马样本来产生数据，或者其他攻击的真实数据。课题希望能尽量贴近真实应用场景，对不同攻击的恶意分布比例没有具体要求。

1、问题：应用层协议

您好，我想确认一下，本次比赛所需处理流量，都是基于应用层协议的吧？

答：不止包括应用层，还包括常见的音视频流量

2、问题：关于算法应该是对不解包的数据包进行处理的疑问
官网发布的视频说算法要快速分析出哪些是正常的业务流量，而对于业务流量之外的可疑流量才需要进行解包分析。这里的意思是说算法应该是对不经过解包的数据包进行处理。这里有几个混淆点：

(1) 官网推荐所使用的 wireshark 所抓到的数据包本身已经经过了解包还原，请问这样的数据包还能否使用呢？

答：使用 wireshark 抓取数据包是可以的，wireshark 抓取数据后还需判断该数据是否是正常的业务流量

(2) 视频中提到的特征提取方法，如可以通过 HTTP 报文中的 user-agent 字段来进行识别，这里是否也已经对 HTTP 报文进行了解包呢？
因此想问一下这里解包的定义是什么？

答：需要对流量数据进行解包，是因为需要通过数据包中的一些特征来识别该流量是否为正常流量。这里的解包指的是对原始数据流的解析。

3、问题：展示内容
有要求以 web 形式或者以插件形式吗？

答：不需要

4、问题：展示内容
功能展示板块可以使用第三方抓包工具体现吗？

答：使用第三方抓包工具，还需要在其上实现对数据的判断

5、问题：关于准确率的问题
这个准确率是指【模型判定为正常流量】和【实际正常流量】之比，还是【模型判定为正常流量并实际为正常流量】和【所有流量】之比，或者说是指其他集合之比，能否细说

答：准确率是指【模型判定为正常流量】和【实际正常流量】之比

6、问题：关于 https 的包

wireshark 抓到的关于 https 的 TCP、TLS 协议的包是否在本题目白流量过滤范围内，我的数据里面是否只能出现 HTTP、DNS、视频流协议的包，对于在 HTTP 之上的协议如 OCSP 此类协议是否需要剔除？

答：本课题可只识别非加密数据

7、问题：功能

产品需要展现实时过滤的功能吗？还是只需要用自己抓取的数据包实验

答：不需要实现实时过滤，但希望能够识别出的白流量类别越多越好

8、问题：关于如何进行流量过滤

HTTP、DNS、视频流量如果特征不尽相同，是否可以分开做，是否只是需要识别应用层的协议包，下面的 TCP、UDP 不用管。视频流量是否能说说主要是通过哪些协议传输。

答：本课题希望能够在模型或系统中对网络流量全部能够识别，因此不止包括应用层协议，还应包括传输层协议。由于不同的应用会使用不同的协议传输，同学们可以通过抓包分析不同的应用。

9、问题：关于恶意流量数据集的使用问题以及机器学习的问题

贵企在之前的回答中说，不能直接使用恶意流量数据集，可是机器学习就一定需要有对比的数据。那么这不就矛盾了吗。还是说我们先用一套简单的规则过滤掉一部分正常流量后，再用机器学习（正常和恶意样本）得到的规则进行进一步过滤？不知道这个问题是否刁钻，可能是我的理解能力不太够。。希望得到贵企回答，感谢！

答：本课题的目的是对白流量进行过滤，因此需要对流量本身进行分析，如果恶意流量数据集也是原始的数据包格式也是可以的。本课题希望使用机器学习的方法能够覆盖所有的流量。

10、问题：数据要求

是否要对抓到的所有报文都进行过滤，包括所有的服务器返回报文，对于数据有没有什么限制，是否完全凭自己的理解来进行操作。

答：对数据没有严格限制，希望能够对所有报文进行过滤

1、问题：如何量化评估模型

因为模型的准确率和模型开销不能兼得，这两方面的权重如何，是否有量化公式

答：最终评估模型的时候，会综合这两方面考虑，但是准确率的权重更高一些，准确率权重占比 70%。

2、问题：关于 https 的包

wireshark 抓到的关于 https 的 TCP、TLS 协议的包是否在本课题白流量过滤范围内，我的数据里面是否只能出现 HTTP、DNS、视频流协议的包,对于在 HTTP 之上的协议如 OCSP 此类协议是否需要剔除？是否要将 https 解密

答：HTTPS 协议的流量也在本课题范围内。HTTPS 不解密。OCSP 类如果占比很高，也需要进行过滤

3、问题：关于解包

上次解答中说到，解包是对原始数据流的解析，就是说，原始二进制比特流吗？赛题中说到希望尽可能做到少解包，也就是说，如果可能，从二进制数据中分析流量特征？

答：可以使用解包的方案。但是直接从原始二进制流识别，可能性能更好一些。

4、问题：关于 tensorflow 以及全流量检测的疑问

1.可以使用 Keras 吗？还是说算法的实现只限于 tensorflow 抑或是需要自己来完成？

答：可以用 Keras 实现。不需要自己写算法。

2.全流量检测是指使用贵司的产品进行全流量检测吗？课题中所描述的全流量检测作为重要的与所设计的模型比较效率的检测手段具体应该如何实现呢？

答：可以先不用关注产品相关的东西，重点关注如何识别过滤处正常的流量。

1、问题：关于 tensorflow 以及全流量检测的疑问

1.网络边界检测设备是指什么？

答：边界检测设备是指流量镜像探针。通过镜像核心交换机的网络流量，检测里面的攻击行为。

2.需要自己设计全流量检测的程序吗？有没有一个确定的性能标准呢？

答：不需要自己设计检测的程序。但是需要提供详细的设计说明文档，规则或者模型。性能标准以最终合入产品的效果为准。

3.具体需要涵盖的白流量有哪些（种类）呢？还是说这里的白流量由学生自己定义（是指不会造成危害或者没有潜在危险的流量吗？）

答：

白流量可以自己定义，因为互联网上网络流量的类别太多，学生可以按照各类流量类型的比例自己定义。

2、问题：请问最终只需要得出白流量过滤的准确率，还是需要还原出过滤后的 pcap 提交呢？

答：最终需要提供详细的设计说明文档，规则或者设计模型，准确率，需要提供部分实验数据。不需要提供过滤后的 pcap。

【A07】高级计划与排产平台【雪浪云】

1.问题：关于作品项目范围的具体问题

题目是“高级计划与排产平台”，但是任务要求部分提及的却是具体调度方面要求。因此我们的问题是此次任务是仅仅聚焦于 APS 调度算法部分还是基于 APS 算法的 ERP 平台

答：

算法和平台均有考虑 更偏向于算法 因为算法有很多约束 会描述的比较丰富 平台的话属于完成整删改查以及报表和插单的设计 这属于比较开放和需要同学们去考虑的东西 不做具体扩展功能的限制 故提及的较少。

2.问题: 可以提供以前的排产表吗
是否有以前公司里面的排产表数据可以提供?

答: 排产所需数据以及结构已提供, 排产结果的话也许会展示。

3.问题: 工序所需前置半成品
工序所需前置半成品 `input_semi_finished_product_id` 是只有装配工序才需要前置半成品还是所有工序都可能需要前置半成品?

答: 提供的数据样例中只有装配工序才需要前置半成品, 实际场景中所有工序都有可能需要前置半成品, 但是可以不考虑。

4.问题: 技术偏重与要求
整体感觉项目偏向于 PC 端, 请问移动端是否需要, 如果需要, 效果要做到什么程度?

答: 工厂中主要以 PC 端为主, PC 端主要以排产逻辑, 调度, 规则制定, 增删改查等, 移动端的话更多只是当天资源/任务的展示。不强制需要移动端。

5.问题: 工序加工和装配过程是否可以中断
工序加工和装配过程是否可以中断?

可以中断是指当天未完成的工序可以在下一个工作日继续开展。

答: 正常生产是不能进行跨天的, 但是一天工作时长可以考虑 8 小时或者不考虑工作时长, 作为新的约束。

6.问题: 机床与人员关系的问题
请问人员控制机床是一对一的关系, 还是说一个人员可以控制两三台机床, 还是说机床前不需要人员呢?

答：人员和机床没有强制对应关系，是根据工序的资源需求情况来的，比如 process_1_a 需要 3 台机床，1 个人员，2 个高级人员，3 台设备。但是 process_1_b 对应关系不同，故不需要关心机床与人员的配对关系。（注：process_1_a 所需所需资源为 3 台机床，3 个人员，其中至少 2 个人员需要拥有高级人员的属性，3 台设备）

7.问题：手动调整是指什么

请问 A07 中说的手动调整是指在排产之后的调整吗，可以具体举个例子吗

答：排产后的调整，现实场景中，计划排产完成后，会因为一些现场的原因如物料不到位或现场临时调整，针对其中的某一道工序的提前或者延后，在手动调整时，需要考虑的是调整后资源，前后置关系等约束是否满足。

8.问题：导入功能

系统必须要有文件导入的功能吗，如果有，以文件的形式导入相关信息的时候，导入的格式是规定了的吗，还我们组自己定？

答：文件格式与数据样例格式相同，需要有基本增删改查功能。

9.问题：关于串行工序

串行工序是指同一时间内一份量的资源只能进行一次工序，但是如果有两份资源的话，就能同时进行两次该工序吗？还是说不管是有一份资源还是两份资源，同一时间都只能进行一次该工序

答：有两份资源就可以同时进行两次该工序。串/并行是定义在工序上的，如 AB 工序串行指的是 A 工序做完后才可以做 B 工序，与资源无关。

10.问题：导入文件的格式是什么

导入文件测试我们的系统的时候，是按照给我们的数据导入，还是有其它固定的文件格式，还是我们可以自己规定

答：文件格式与数据样例格式相同。

1、问题：所给数据出现乱码

给的 Excel 里面中文部分都是乱码

答：提供的 csv 文件是以 utf-8 编码的格式，windows 的 excel 直接打开会有乱码问题。可以百度查询一下 excel 打开 csv 的乱码问题。例：打开 Excel，执行“数据”->“自文本”，选择 CSV 文件，出现文本导入向导，选择“分隔符号”，下一步，勾选“逗号”，选择 utf-8 的编码下一步，完成。

2、问题：关于物料约束问题

再给出的样例数据以及上次答疑的问题中，都未提及物料的种类（除半成品之外的原材料等）、存量，所以该计划中是不需要考虑物料约束以及库存量这两项吗

答：物料以及库存量在该计划中不需要考虑，默认物料和库存都是满足需求的，弱化约束。

3、问题：在排产过程中，需要考虑 setup 时间吗

在生产中切换不同的产品进行生产，需要考虑中间切换过程的时间吗。如果考虑，是会给出具体的数值，还是暴力认为尽可能把相同工艺路线的生产进行合并一起生产？如果不考虑，那是否只要满足交货期并且 make-span 最小就可以，订单加工的顺序无所谓？

答：不要求考虑切换过程的时间，弱化约束。

4、问题：数据 aps_order 中订单问题

数据 aps_order 中订单的时间跨度一年多是否有误

答：order 中的最晚交期为随机生成，实际不成为约束，在本次计划中弱化了该约束，主要还是优化 makespan。

5、问题：关于样例数据的完备性与数据库设计问题

1.在给出的样例数据中所有数据是完备的吗？比如 aps_resource.csv 中，人员_A 既隶属于加工车间又属于装配车间的情况是否为真实情况；资源类型仅涉及机床、人员、设备，对应的 resource_attributes 为机床、高级机床；人员、高级人员这种，故问做排产建模与算法时，仅需要考虑所提供的资源约束吗？还是说需要考虑资源的扩展性，比如机床中又添加了高级机床 B，或

者其他诸如模具之类的约束。

答：在 `aps_resource` 表中，`name` 不是唯一标识，`resource_id` 是唯一标识，不同的 `resource_id` 为不同的具体资源；

资源类型与资源属性的区别为，资源类型是某一个资源所属的最大类，资源属性为某一个资源所属的所有类，`basic_attribute` 指的是某一个资源所属的最小类，关系为：高级机床也是一种机床资源，是向上满足的，即如果某一工序的资源需求为机床 2 台，可以是 2 台机床，1 台机床 1 台高级机床或 2 台高级机床；

资源的扩展性不做强制要求，但可以作为功能的鲁棒性评分。

2. 关键数据库表是需要完全按照给出的样例数据设计，还是说可以根据赛题和系统设计进行适当的拆分与增添？（就和问题 1 有关，比如拆分为机床和设备表，但给的数据不完备的话就很难考虑）；

答：可以进行适当的拆分和增添，但是我认为机床和设备都是作为生产资源作为管理，从生产资源角度统一在一张表上是较为合理的。

3. 数据库字段名和对应的数据类型是否以给出 csv 文件中类型对应？比如 `product_id=product_1`，依照案例的理解这就是个 `char` 类型的。

答：是的，`id` 列是 `string` 类型。

6. 问题：关于显示的分辨率设置

Web 端是需要适配超大屏还是仅供普通的 PC 使用

答：以普通 PC 为标准。

7. 问题：资源类别以及生产车间是否只有官方数据中的那几种？

资源类别以及生产车间是否允许用户输入其他值呢？还是说用户只能从数据中的那几种中进行选择？

答：资源的扩展性不做强制要求，但可以作为功能的鲁棒性评分。

8. 问题：关于报表和插单

看到答疑里有说平台需要完成报表和插单的设计，请问报表具体需要展示哪些内容？另外，插单是指实时添加新订单吗？是否意味着排产结果是动态变化的？

答：报表的设计主要是想体现功能的展示性，具体内容不做强制要求；
插单是实时添加新的订单，如果考虑插单即排产结果是动态变化的。

1、问题：比赛作品是否需要根据雪浪云的数据进行实验
比赛作品是否需要根据雪浪云的数据进行实验

答：需要 至少需要保证其排产基本约束的满足 如前后置，资源上限约束。

2、问题：排产算法
能否提供一些排产算法的推荐方法或者相关实现代码

答：启发式的方式推荐 ortools；mip 的方式可参考开源求解器 cbc；遗传算法等可自行搜索对应解法。

【A08】归家行动-智能救援指挥系统【融创软通】

1.问题：调用百度或者阿里云平台
人脸识别以及其余识别是否可以借助百度以及阿里的相关 API 实现

答：可以借助其他相关 API，以能实现功能为主即可

2.问题：文档具体要求的疑问

1.技术要求与指标中的第七点“实现手机控大屏”是什么意思？

答：用手机能操控可视化大屏（比如电视大屏）的数据变化

3.问题：A08 项目作品移动端是否需要实现多平台
作品的移动端实现能否只实现 Android 端

答：仅实现 Android 端就行

4.问题：技术要求与指标【4】轨迹跟踪请问是指什么意思？

轨迹跟踪是用于跟踪走失老人的吗？可是我们怎么能做到跟踪走失老人的轨迹呢，如果能够知道老人的轨迹，那就直接能找到老人了呀。

答：老人没有 GPS，记录救援者的轨迹

5.问题：技术要求与指标【4】自动报警是是需要干些什么？是指需要在客户端上放置报警电话吗？

答：自动预警，如果找到或找到类似的目标（80%以上）需要 App 进行提示

6.问题：技术指标和甄别算法的【轨迹跟踪】具体指什么？

轨迹跟踪是指老人身上自带的 GPS，根据 GPS 的数据分析老人的是否有异常轨迹吗？

答：老人没有 GPS，记录救援者的轨迹

7.问题：技术指标中【自动报警】具体指什么？对什么的自动报警？

答：找到老人，或相似度较高的老人，app 进行主动提醒

8.问题：技术指标的【行为识别】是指什么？是指步态识别吗？

答：根据老人的行为和动作进行判断或根据老人的步态进行识别判断。

9.问题：百度 PaddlePaddle 人工智能平台属于【利用 baidu 或 aliyun 的 AI 数据接口构建深度学习平台】吗？如果不是，能否举例说明一下【AI 数据接口构建深度学习平台】这一概念

答：利用相关公司的 AI 数据接口构建深度学习平台，如百度，阿里等

10.问题：关于手机控大屏的问题

手机控大屏是指监控平台的图表可以全屏展示吗？类似于最小化、最大化窗口那种？

答：比如数据图表较多存在多屏显示的情况，用 APP 小屏可以控制大屏的数据，可以翻页。

11.问题：人脸识别

是可以直接调用百度、阿里云的人脸识别 API 吗

答：可以

12.问题：关于反馈信息通过短信回传到系统的问题

这是不是要我们实现一个短信回复和发送的功能？但是这个接口功能很浪费钱，所以反馈信息能不能不通过短信回复了，发一个链接，让他通过微信小程序或手机 app 提交？

答：可以

13.问题：请问题目中的手机控大屏是什么意思
请问题目中的手机控大屏是什么意思

答：用 APP 小屏可以控制大屏的数据，可以翻页。

14.问题：救援工作岗位是不是不用理？
救援工作岗位有很多，仅需实现合理找人的流程，是不是说不用为队员增加工作岗位这一栏信息？

答：自选

15.问题：家属更新的信息由队员填还是家属填？
家属更新的信息是由队员填还是家属填？

答：都可以

16.问题：自动报警
题目中所给的要求自动报警是什么意思？在什么情况下触发报警，是信息上报到公安局还是系统发出劲爆提醒使用者的意思？

答：不是上报公安局，只由系统发出劲爆提醒使用者即可

17.问题：手机控大屏
题目中所给出的手机控大屏是什么意思？是说用手机端来控制系统还是有别的什么意思？

答：大屏指监控显示屏，不方便操作，用手机进行控制数据变化或翻页等等

18.问题：关于甄别算法研究问题
利用 baidu 或 aliyun 的 AI 数据接口构建深度学习平台，是指用百度的 AI 接口去实现吗？还是用百度深度学习平台训练好模型，再用 openCV 实现？

答：用 AI 平台，并训练模型，实现方式不限。完成目标即可。

19.问题：关于志愿者组队的机制问题

请问每个队伍人数是否固定？一个任务是否仅由一个队伍完成？

答：队伍人数不固定，一个任务可以由一个队伍完成

20.问题：关于移动客户端的问题

项目移动客户端部署是要求在 Android, ios, 微信小程序, web 页面都要实现部署，还是任选其中一种合理方式部署？

答：任选一种，方便即可，不要给使用者增加难度

21.问题：请问微信小程序、安卓、IOS 等端口都需要完成开发吗？

题中说“队员：能够便捷的在多平台（Android、iOS、微信小应用程序）进行安装部署并使用。”，是需要完成所有端的开发吗？

答：完成一个平台即可，自选，技术不限。

22.问题：请问归家行动启动判定标准是如何制定的？

企业提供归家行动启动判定标准吗？还是说后面可以直接在系统中再添加一些标准？具体是怎么样的呢？

答：其他不提供判定标准，自行判定启动判定标准，合理即可。

23.问题：第三方成熟商业设备

第三方成熟商业设备是什么？？与手机控大屏相关吗？？

答：比如有第三方的摄像头，定位系统等等可以使用，可以利用第三方硬件。

1、问题：对于短信服务

从 2020 年 12 月起，阿里云和腾讯等都不支持对于个人短信通知的服务了，请问对于我们参赛选手，提交该方案的实现计划（从企业角度），不实现相关代码是否会有影响评分？

答：如果没有短信服务，可以考虑使用 微信公众号 来完成类似问题，亦或者是使用电子

邮件模拟该功能，也可以在文档中指明预留了 对应的接口以便后续对接使用，比如，只
要选择对应平台，并填写 key 值的内容即可启用平台的短信提醒功能。

2、问题：步态算法问题

步态算法对数据的要求比较苛刻，家属一般无走失老人走路的视频，怎么建
立数据库进行比对呢？

答：该问题在上次已经做了说明，这不是这个平台要求的功能，而是作为独立的研究部分，
并不要与现在的平台整合到一起，对于我们来说，评判的依据包括有数据支撑的论文，独
立的验证项目等等均可

3、问题：后端

后端功能组件编程是否限定语言

答：语言没有限定，Python、Java、.net 均可以，我们推荐基于 Java 的 Springboot 作为开
发框架，但不是绝对的

4、问题：请问会出现被分配到同一任务的几个搜救队员间互不认识的情况 吗

不是很了解救援队内部组织是什么样的，所以想请问一下会不会几个搜救队
员间互不认识导致无法相联系，谢谢！

答：这种情况完全不存在，上次已经说明，xxx 救援队（软件目标用户）是一个专业化较
高的救援队，其志愿者进入需要一定的门槛，而且他们经常有配合活动，且人数并不是很
多（团队大概 30 人左右），因此全员可以说都互相认识

5、问题：对于赛题描述和答疑不一致引起的矛盾问题

如果赛题文档描述和视频答疑或者文字答疑有出入，以什么描述为标准？

答：请以赛题文档为准，因为那个是官方发布的公示文档，具有权威性。

6、问题：是否需要做公众号？

录播中多次提到了关于信息的通知和公众号，我们是否还需要做公众号呢？还是只需要实现和公众号的连接？

答：开发阶段对于公众号的功能要求不是必须，但如果做微信小应用程序则必然有公众号，这个是躲不开的，只是对于公众号的功能和内容不做要求而已

7、问题：三个端

请问移动客户端、后台管理端、可视化监控端是要开发三个独立的端吗

答：移动端和后台管理端一定是分开的两个项目，而可视化监控可以与后台管理端整合在一起，也可单独独立出来

8、问题：志愿者组织的组织架构是什么样子的？

您好，我们想知道搜救志愿者的组织架构是怎么样的。比如是否有救援指挥中心以及前线总指挥部？或者只是一个地区由一支队伍全权负责管理？是否有总队长、小队长、普通队员的区别，每次搜救时，搜救组织的小队长是相对固定还是临时指派？系统是否需要考虑全国省、市、县分级的组织架构？

答：当前该救援队对于这个项目的要求仅限于在特定地区（例如天津范围内）使用，他们有专门的架构，但在协助找寻阿尔兹海默症老人这件事上，所有的队员都会被视为平等的节点，仅有一个或两个（可由平台管理者，例如总队长指定）可以访问管理端，进行例如搜救任务级别调整、关闭任务、添加认证队员等必要的管理操作。其他人都仅有客户端的访问权限，因此在该项目范围内，可暂不用考虑救援队的组织架构。

1、问题：关于安装配置说明文档

企业要求提交的材料中，安装配置说明文档指的是数据库在服务端的安装配置吗，具体内容是数据库里各个业务所需要的表创建吗，提供一个 sql 脚本？

答：关于问题中提到的数据库问题，公司这边仅需要提交数据库的导出脚本（.sql 文件）即可，并说明所用数据库的版本即可

【A09】轨道交通智慧客流分析预测【八维通】

1.请问能否告知所给数据的具体来源呢？在企业方所给的任务清单中提到了增加温度、天气的影响因素，也提到了制作热力图用于展示，然而这些都依赖于具体的城市及地铁结构信息。如果不能给出具体城市的话，我们是无法找到与数据相匹配的天气信息的，“为使模型跟贴近实际，团队可以自己模拟加上相关的影响因子例如：温度、天气等等”这一点故而不可能达成。同样的，如果不能得知具体的线路拓扑结构，仅靠现有数据推算的话，是不能保证得到的拓扑结构的正确性的，线路结构不正确，模型准确性也就无从谈起；而且如果由我们自己根据数据推算线路结构的话，我们只能知道站点间的连接情况，而无法得知真实的地图，那么企业方要求的热力图也是不可能准确地画出来的。这些问题，希望企业方予以考虑。

答：

关于数据来源以及城市的话，涉及到保密不便透露。因此给出的用户 ID/站点名称都是修改过的，但是所有行程数据都是实际的。因此，线路拓扑图各个团队可以自己设计，这个不影响热力图的展现。如示例图



2、有老师反映官给的数据有误，如钱不对劲（有非整数的钱），trips 里面有几个站台在 station 里面没有（5,155 号）

答：

test_trips 表里面有近 80W 行数据，1. 其中里面有 77 条金额不是整数的，可以去除，或者自行修改，并不会影响最终计算结果。2. 有 32583 条行程里面的 in/out_station_name 不在 station 基础表里面，这个是由于原数据的差异导致的。按照比例看来，异常数据只占全部数据的 3.7% 因此不影响结果，可以不予理会或者抛弃。同时作为大数据处理，数据的清洗和容错是考虑的一方面。

3.问题：有关 A09 赛题数据存在一些纰漏的问题

请问还有更多数据吗？目前所给的行程数据只有 2020 年 4 月到 7 月的，这些数据是显然不够的。每个月客流量情况影响因素都各有不同，想要以前一个月的数据准确预测出下一个月的流量情况显然是不科学的，要得到准确的模型，我们在训练时需要进行年际间的纵向对比；加之受疫情这个黑天鹅事件的影响，今年的数据必然是相对特殊的，以今年的数据进行训练，得出的结果也显然无法代表普遍情况。因此我们希望能够得到至少近三年的完整数据。

答：全部提供的话将是海量数据，基于客观条件无法提供，希望团队能够通过给定的数据基础之上进行设计工作。同时，如果数据不满足的话，团队也可以自行模拟数据，只要最终算法和设计在逻辑上是合理的即可。

4.问题：题目提供数据所对应地图

请问本次题目所提供的数据是否有数据所对应的地图？

答：考虑提供，需要时间准备

5.问题：我想咨询下，该题的地铁网络拓扑结构会给吗？或者说给一个站点之间关系的数据集自己来搭建地铁网络，如果没有网络的拓扑结构很难根据一个乘客的入站和出站点去推测其旅程路线。

答：考虑提供，需要时间准备

6.问题：赛题数据使用

除了企业给的数据以外，可以使用自己抓取的数据吗？

答：可以自行抓取并且使用

7.问题：对于表格字段的意义不理解

(1) station.csv 表格中的编号是指站点的编号吗?是否意味着相邻编号既为相邻站点?

答：考虑提供拓扑图

(2) station.csv 中的行政区域是二级行政区吗?user.csv 中的区域是三级行政区吗?两者之间的联系是什么(以什么形式包含?)?

答：station 表的指的是市里面的区，比如杭州市的上城区、下城区、滨江区、西湖区等。User 表中的为身份证中的中华人民共和国行政区域代码，比如杭州市、烟台市等，具体可以在网上查询

(3) trips.csv 中的渠道编号的含义是什么?

答：系统其他内部含义、可以不用理会

8.问题：请问客流预测的时间间隔是分钟、小时还是天呢?
请问客流预测的时间间隔是分钟、小时还是天呢?

答：预测某一天的客流即可，如果能够做到小时则更佳

9.问题：地铁计费规则

可不可以公布一下计费规则，因为存在同站进出的 OD 收费标准不一样。我们是按照时间和距离结合收费吗，如果是的话可不可以公布一下详细的计费规则!

答：同站进出按照 2 元计费、最高票价 6 元。地铁实际计费规则过于复杂，建议阶梯票价可以按照站点个数来进行，阶梯可以自由定义比如 5 站之内 2 元，10 站之内 3 元，15 站 4 元，20 站 5 元，最高 6 元封顶。

9.问题：数据问题

为什么数据里会有这么多同站进同站出的人呢

答：实际数据如此，总共不到 7000 条左右，也有可能是地铁员工每日巡检产生的线上测试数据，不影响结果。

10.问题：关于天气温度等因素对客流影响的数据如何获取以及关于车站表站命名的疑问

如果要用天气温度作为影响因子去构建预测模型，这个温度的数据如何获取呢？是自己去做一个温度信息的数据集吗？这个数据集是只作为展示效果呢还是必须要真实可靠？

如果是从网络上爬取天气数据的话方便透露是哪一带的天气情况吗？

答：天气因子可以自己抓取，考虑西南一带的。天气因子作为一个参考量，可以看是否对客流有影响。

附：对于站点信息的数据表，发现没有站点的路线是相同的，基于数据是真实的，那么是不是换乘站在不同路线会标注不同的站名。如果是这样的话，自己设计拓扑图的时候是不是自己设计几个换乘站呢？

答：考虑提供脱敏的信息，需要些时间准备

11.问题：关于大数据实时计算能力

请问这里的大数据实时计算能力是否意味着，在已给出 80 万条数据外，后期需要提供接口接收新的行程数据加入模型进行分析

答：新数据基于数据库的方式即可，无要求具备接口或者实时消息接收能力。

12.问题：你好，请问“站点 OD 客流量分析”具体应该怎么理解？
你好，请问“站点 OD 客流量分析”具体应该怎么理解？谢谢~

答：OD (Origin Destination) 的意思。OD 匹配指用户的进出站匹配。本数据中在 trips 已经给出匹配好的行程。

13.问题：有关拓扑图的问题

想问一下各位老师，这个赛题的拓扑图是必须要实现的吗？

需要考虑每次行程除进出站点外经过的站点吗？

答：只需要考虑进出站即可，中间途径那些站点不用考虑。

14.问题：关于数据的时间

我们团队发现所给的数据节假日和 2020 年对不上，而跟 2019 年对的上，请问这个是主办方提供的数据出错还是按照这个数据继续进行？

答：数据包含 2019 年和 2020 年的。如果节假日对不上的话，可以自行考虑先修正。近期提供更新数据。

15.问题：关于实时计算

是否需要使用分布式存储与分布式计算？实时计算间隔时间需要精确到多少？

答：按照各团队的能力自主设计分布式计算以及实现方式，计算间隔不做硬性要求，按照团队能力建设即可

16.问题：路网信息图是否需要自己推算得出

由于 station 中的站点名是没有重复的，所以难以推测出线路之间的交汇换乘点，也难以推测出站点在线路上的分布顺序，这样的话，就难以得出路网信息图。

答：考虑提供

17.问题：数据集中存在的异常情况，在计算的时候是选择抛弃还是考虑进去

1.数据集中存在某个站点在某一月份的进站人数为 0 的情况，如 sta89 在 2 月份的进站总人数为 0，这种情况是由于该站点关闭还是其他原因造成的，计算的时候需要考虑进去吗？

答：可以作为异常数据点

2.数据集中存在六千多条出站与入站为同一站点的行程记录，这是属于系统异常数据还是正常的用户行为，因为从正常的出行角度考虑，出发地与目的地是不相同的，若是相同的则难以推演用户的具体路径以及行为特征。

答：真实数据，但是无法判断实际用户情况，很大可能是地铁员工做每日巡检产生的一些线上测试数据，不到 7000 条此类数据，整体上不影响数据分析

18.问题：企业提供的数据内容

- 1、提供的数据 workdays2020.csv 文件中，第 115 条记录，20200424 这一天的节假日属性是 1x，请问是这是数据提供错误吗？
- 2、提供的 trips 行程信息中存在 userid，在 user 用户信息中找不到匹配项，这是否是数据不一致，还是 trips 中的 userid 无需参照 users 中的 userid
- 3、trips 行程信息中，有的站点在 station 信息表中没有对应，例如 Sta5，在行程信息中有，但是在站点信息中没有

答：1. 请改正，谢谢。2. Trips 和 user 表并非完全同步，user 表有含有历史的用户信息，因此，存在 user 表中的用户没有行程的情况，这个没有关系，只要 trips 表中的用户都能在 user 表中找到即可。3. 考虑做些数据修正

19.问题：数据问题

给定数据集没有换乘站点信息，无法构建地铁网络拓扑模型，不好建立各线路间的联系，也难以展示流量分布。企业是否可以补充相关信息？

答：考虑提供，需要些时间准备

20.问题：关于轨道交通的各个站点的拓扑数据问题

你好，在官方给出的数据中，仅仅包含各个站点是属于哪条线路，而我们更希望官方给出各个站点之间是如何相连的数据。要求给出此数据并非为了数据展示，而是在数据预测上能够提高精准度。此数据并不会涉及轨道交通数据的隐私性。

例如：

(1, 3) 表示站点 1 和站点 3 是相连

(2, 8)

.....

21.问题：关于数据
请问有客户在地铁站中转线的相关信息吗？

答：无，只能靠最优路径进行计算

22.问题：关于数据集对应的轨交线网
对于官方所给的数据集，不知道能否给出具体的轨交线网，不必透露具体的信息。

要进行线路断面分析的话，势必要分上行客流和下行客流，这个需要知道站点之间的相对位置。

数据集当中对站点的编号是没有重复的，但是换乘站是势必存在的，应该是数据脱敏造成的。

不知道可否给出一个具体的线网图，或者是给出能够确定整个线网的提示。

谢谢！

答：考虑提供

23.问题：提供的数据集中行程数据是否有增删
根据我们的数据分析发现客流量情况与实际有差距，比如 2020 年 6 月 14 日 18 时-19 时 163 个站点的进站总人数为 495 人次，平均每个站点在 18 时-19 时进站 3 人次，与实际地铁运营情况并不相符。所以提供的数据集是否是删减了一定比例的，在这种情况下我们是自己模拟填补数据还是仅根据已有数据集训练，训练过程中我们需要完全遵循该数据集吗，可以使用别的数据集吗

答：数据样本只取了某几个票种的乘车行程（全部行程的话将是海量数据，无法提供），故可能有些站点的行程看起来不正常或者差别较大。团队可以根据自己的设计思路以及理

解决采用或者不采用某些认为是异常点的数据。新增数据集也可以但是需要说明新增数据的依据。

1、问题：赛题对于预测的精度有要求吗？

在任务清单中，要求（2）：建立准确的预测模型，对以上全部或者若干项进行分析并获取结果。

请问赛题对于预测的精度有没有一个具体的要求？有没有误差的限制？

答：没有特定的精度要求。依据各个团队自身的能力建立模型并能做出解释。

2、问题：给出的数据量不足在建立预测模型时遇到困难

示例数据的行程数据（trips.csv）约为 80w 条，将近七个月的数据，一共 163 个站点（station.csv），平均到每个站点每天只有 23 条数据，这样在建立预测模型时遇到困难

答：站点行程分布并不平均，并且同一站点每日分布的行程也不平均。这个是实际的数据，在行程的站点分布以及时间分布还是有代表性的。建议按照实际先做计算。

3、问题：关于线路站点的顺序

能否提供各个线路的站点的顺序，即起点和终点，以及中间站的顺序

答：新提供的 test_station 里面包含了每个站点的顺序，编号顺序为线路的上行方向（序号由小到大表示站点从西到东，从南到北排列顺序）

4、问题：关于技术

请问可以使用 echarts 吗？

答：可以，不限任何前端技术实现方式

5、问题：关于数据

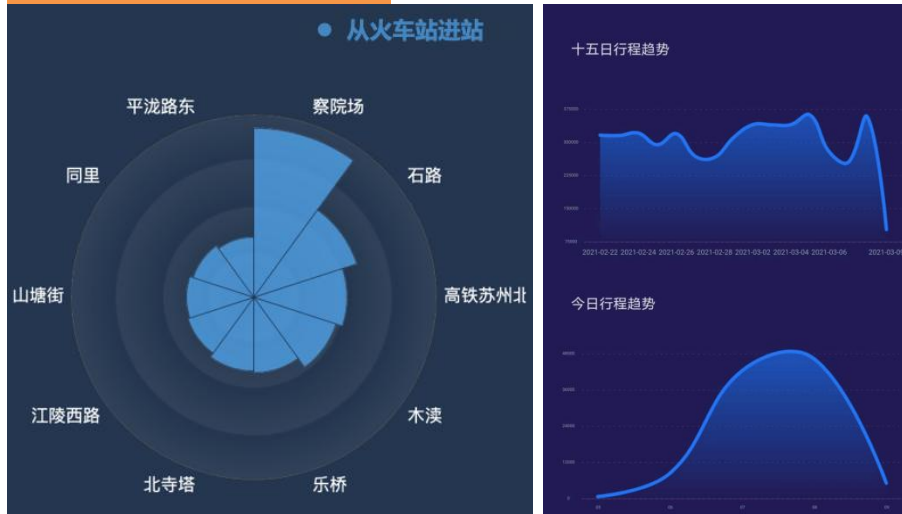
请问我们可以自己找数据吗？因为那个位置信息太不明确了

答：不是很清楚自己找数据是什么意思，是自己模拟全部数据吗？如果数据有缺失或者在实现某些算法模式方面数据有误差，团队可以自行添加合理的数据作为补充。

7、问题：关于客流统计的相关问题

任务清单中 (1) 要求统计客流，但除了任务 c) 外均未说明用何种统计方式统计，请问是用各地地铁采用较多的客运量统计方法，还是集散量或者乘降量

答：按照进站出站客流分析



8、问题：预警评判标准

请问预警的评判标具体是什么？因为如果根据客流量的话涉及地铁的运力，拥挤度的计算也跟车站的面积有关，能否提供相关数据

答：开放式设定预警标准，标准值各个团队请自己定义，并且说明推断数据即可。例如拥堵度可以简单设定一个假定值就可以。如果需要准确定义拥堵度，客流密度等方面的精准定义的话，那么又涉及到客流流量、速度、密度、站内设施布局、干扰因素、通行能力等诸多因素，这些都属于另外的研究课题了。当然有可能的话，也欢迎各个团队能够提出自己的拥堵度模型或者客流模型。

9、问题：赛题数据没有给节点站点的拓扑图结构

赛题数据没有给节点站点的拓扑图结构，我们可以根据赛题给的站点和区域数据自行构建合理的拓扑图信息吗

答：如问题 3，站点方向编号顺序已经给出在新的 test_station 表里面。站点对应的区域也有对应关系。

11、问题：数据提供问题以及线路图提供相关问题

1.我们在数据清洗过程中发现，6 元的最高票价是否有误？因为有大量票价 7 元的数据，和 6 元几乎一样多。7 元以上的票价就只有几百条数据，像是异常数据，这种情况的话就想问下此题目对应城市的真实最高票价到底是多少呢？应该是 7 元吧？

答：正常票价规则是：最低 2 元，最高封顶票价 10 元。在实行优惠票价的时候，最高封顶 7 元

2.希望提供具体的发车时刻表，对设计最短路线能起到一定帮助

答：目前没有每个站点的具体列车时刻表，但是可以按照早上 6:30 发首班车，晚上 10:30 发最后一班末尾车计算。每个站点的间隔时间可以认为是 2 分钟。

3.答疑文档里提到了会更新提供的数据集，想问下具体更新的是哪部分呢？是否影响当前比赛进程

答：test_station 已经给出，编号顺序为线路的上行方向（序号由小大到表示站点从西到东，从南到北排列顺序）

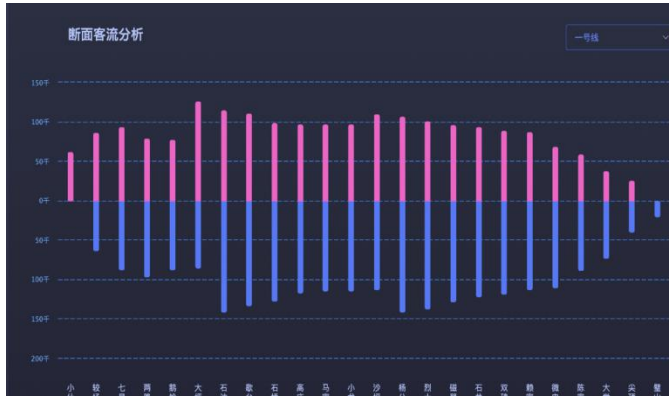
4.同样也是答疑文档中说的会提供线路图，能否尽快提供，最快什么时候可以提供

答：请参考 test_station 中的站点序号理解站点之间的相互关系。线路图建议团队自由绘制

12、问题：线路断面（按站点）流量

请问这个是怎么理解，是各线路在两个站点之间的客流量吗

答：这个专用术语可以自行网上搜索解释。简单来说就是一条线路在某一个时刻一个方向上的客流量。以下为附图可作为参考。



14、问题：线路中的站点顺序问题

题目未直接给出同一线路中各站点的相邻关系。同一线路中站点的顺序是否与 station 文件中所给的编号一致?例:在 station 文件中，将线路 1 中所有站点按照'编号' 排序，得到的结果是否为各站点的相邻关系?或是否可以直接给出线路拓扑图?

答：请参考新的 tes_station 编号顺序为线路的上行方向（序号由小大到表示站点从西到东，从南到北排列顺序）

15、问题：未给出不同线路间转线的站点

既未给出同一线路的运行顺序，又未给出不同线路间转线的站点，如此，无法通过乘客的进站与出站站点预测其行程路线。若官网无法提供拓扑图，请求提供各线路的站点顺序（即站点相邻顺序）与不同线路间的转线站点

答：请参考新给出的 test_station 了解站点之间的关系与顺序。编号顺序为线路的上行方向（序号由小大到表示站点从西到东，从南到北排列顺序）

1、问题：天气信息

由于数据脱敏，自己寻找天气数据和该数据城市不吻合怎么办

答：天气数据见附录 excel 文件以供参考（官网及群文件中下载《城市天气数据-【A09】轨道交通智慧客流分析预测【八维通】》）。

2、问题：要求提交的分析模型具体是指对算法的分析，还是对客流的分析模型？

官方文件中要求提交分析模型，想知道此处的分析模型是指对算法的分析，还是对客流的分析模型

答：客流分析算法。

3、问题：客流的定义

请问对于客流的定义是进出站人数全部算上还是说只单纯统计进站的人数

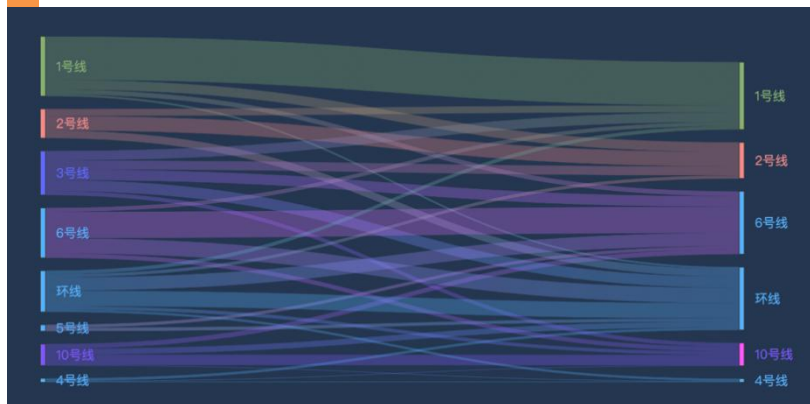
答：按照进站客流或者出站客流统计即可

4、问题：关于新的数据

请问那个新的数据没有一个是中转站吗？168 个站点 16 条数据，12 条线路是平行的吗？

答：目前的数据中不含中转站点信息，但是不影响本题目要求完成的内容（不要求计算最优换乘路线，最短换乘路线），另外基于当前数据仍然可以计算线路换乘流量分析。例如

：



1、问题：线路交汇点与换乘点提供问题

我们需要知道每条线路之间的交汇点，以及单条线路可能存在的分叉点信息。第三次答疑更新的信息只告诉了我们单条线路的站点顺序，然而仅仅是这

样的信息通过数据我们完全可以计算得到（事实上在答疑之前也已经算出来了），因此这对我们来说没有任何作用，在帮助我们了解线路结构方面也没有丝毫作用。是否可以提供更有效的数据？

答：

1. 由于元数据限制的原因，数据表不提供换乘站点数据（即数据中没有不同线路出现相同站点名称的情况）。
2. 不要求计算换乘分析，因此缺失换乘站点不影响。
3. 参考的拓扑图中为了表示线路之间的关系，有出现线路交叉以及交叉上存在的站点，但是该站点并非实际换乘站点，但是如果团队需要参考该拓扑图，并且需要视交叉线路上的站点作为换乘站，并且由此做相应的拓展分析的话（如换乘、最优路径），可以自行假定换乘站点并且修改必要的数据库。
4. 拓扑图仅做参考使用

2、问题：关于新增数据 id 相同的问题

请问给出的新表当中出现了 id 相同的情况，是考虑为换乘站吗？但是如果只有这五个站是换乘站明显是不够的，其中三号线和十二号线有三个相同 id 的情况，可以解答一下吗？

答：元数据里面 id 有重复，不过本数据中 ID 没有业务任何含义，因此可以不用考虑重复 ID 的事情。并不影响统计分析结果。

3、问题：数据解释问题

请问 0309 新上传的那个 station 文件中最后一列 sequence 是什么意思？

答：sequence 表示的是该站点在该条线路中的位置，也就是站点序号从左到右，自下而上依次增大。

4、问题：站点位置关系

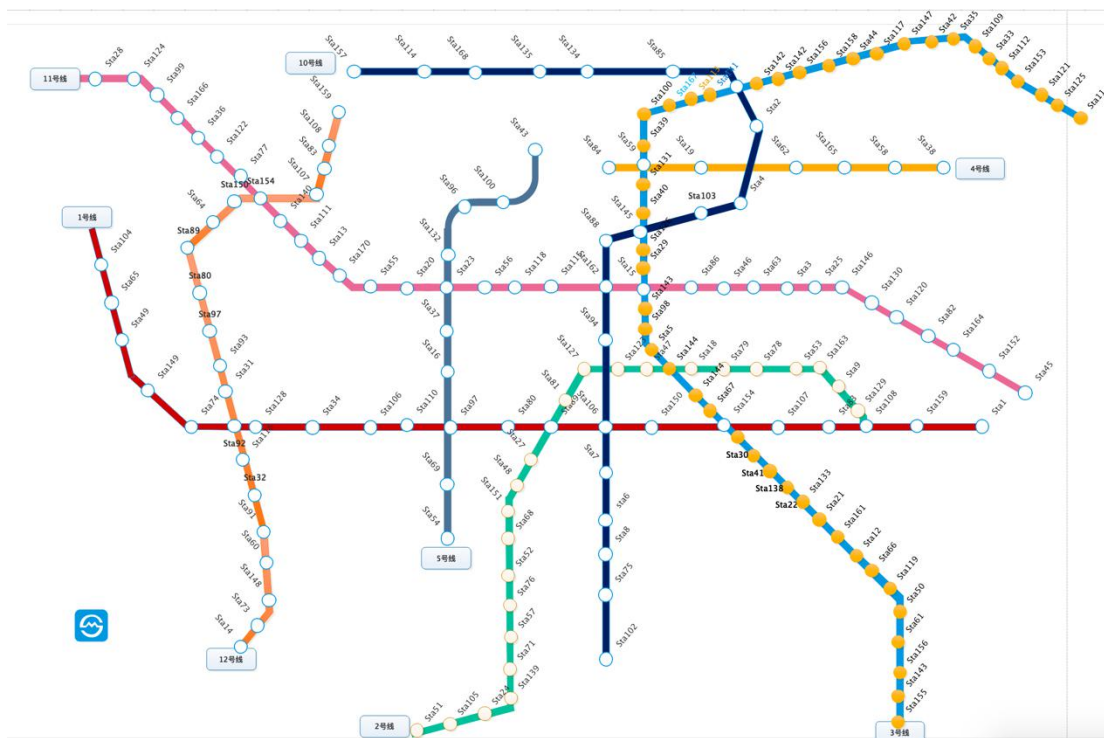
换乘站是由两条线路上的不同名字的两个站组成的吗，这两个站所处的地区是否为同一个

答：

1. 由于元数据限制的原因，数据表不提供换乘站点数据（即数据中没有不同线路出现相同站点名称的情况）。
2. 本题目不要求计算换乘分析，因此缺失换乘站点不影响。
3. 参考的拓扑图中为了表示线路之间的关系，有出现线路交叉以及交叉上存在的站点，但是该站点并非实际换乘站点，但是如果团队需要参考该拓扑图，并且需要视交叉线路上的站点作为换乘站，并且由此做相应的拓展分析的话（如换乘、最优路径），可以自行假定换乘站点并且修改必要的数据库。

附：参考拓扑图，

1. 仅供描述线路之间的大致关系，站点和线路的关系符合 test_station 表中的站点和线路之间的关系，包括站点的顺序和 test_station 中的 sequence 相吻合
2. Test_station 表中由于元数据的限制，不体现换乘站点，所有站点仅属于一条线路
3. 拓扑图中的交汇站点仅做示意，如果团队需要视该拓扑图上的交汇站点为换乘站点，所需的数据修改及调整请自行完成。



【A10】教学视频语音提取文本系统 TEFS 【文思海辉】

1.问题：对部署环境的疑问

假定当前部署环境选定为麒麟操作系统，贵公司对软件部署环境的要求使用的是麒麟桌面操作系统还是麒麟服务器操作系统，在参赛团队无法获得麒麟服务器操作系统的情况下，是否一定需要用麒麟桌面操作系统（优麒麟）运行后端服务？

答：不需要，但是需要用同核心的系统:银河麒麟 对应 Ubuntu ，中标麒麟对应 Centos ，不要最新版

2.问题：视频接口

我们团队想要一下视频的来源

视频我们存在了 96K 上，本地存在了一个文件夹里：

答：视频在我们这里时存放在了 项目目录/upload/temp 和 /upload/courseKpoint/video/四位数字年两位数月两位数日 (20210202) /目录下, 时 UUID 产生的 .mp4 文件

3.问题：文本替换是否需要录音

在出题方提供的视频中提到，需要老师对文本进行修改，再将修改后的文本转为语音拼接回去，而在出题文档中则提到需要提供录音功能以替换文本，所以老师是对要修改文本段进行重新录音再由系统拼接语音，还是老师仅进行文本修改由系统生成语音并拼接。

答：文本能转语音最佳，为了预防难度过大，最保守方案是录某一段语音。期望是文本转语音。

4.问题：文本再向语音转换过程的问题

您好，文本修正后再向语音进行转换，是否是去除视频中讲师说的那一句话，使用修改后的话进行代替；声音就是文本转化成语音后的电脑发声吗？

答：是的，这个语音模型好好训练哦，实在不行，就用老师录的吧

5.问题：对于教师是否需要提供审计功能

对于教师是否需要提供审计功能，如要提供是否需要与管理员有所区别，如无法查看其他教师相关编辑信息。

答：审计功能是指针对某个教师的音视频制作以及编辑情况的汇总，如某个老师一共有多少个视频上传成功，编辑修改了多少个视频，每个视频中被编辑修改的时间长度以及具体位置

6.问题：需不需要写一个新的网页

请问我们这个项目是还需要自己写一个网页还是只需要在给的文思海辉自主学习平台网页中继续完善需要的功能就可以了？

答：自己设计网页，加入平台仅需要一个链接即可。

7.问题：关于修改文本后的操作

修改文本后，是将文本转换成语音，再将语音和视频合并吗，如果是这样转换成的语音和老师声音出入太大，所以修改文本之后的具体操作是文本转语

音，还是录一段语音与正确的文本语音流合并？如果是后者，就应该不需要修改文本了，只需重录错误的语音部分

答：文本转语音是核心，是期望，这里需要调节声音模型，实在无法做到，那就采用重录，但是重录的内容也是按照修改的文本来读，所以需要修改文本。

8.问题：关于视频流处理的工具问题

能否选取别的工具或方法处理视频流？（赛题介绍视频里提到的是 sphinx4，但好像不是强制要求）

答：sphinx4 是处理声音模型较好的开源框架，如果您有更好的选择，前提是不能使用商业化破解软件，试用版也行，能和代码结合在一起，即：用代码解决问题。实在做不到，可以考虑工具，只不过效果差一些。

9.问题：关于补充资料压缩包中的数据库的问题

1. 官方提供的补充资料中的两张表是否都不可修改，且都必须包含在系统中使用；

2. 如果不可修改，能否提供更为详实的说明，亦或者可以按照团队的意愿自行理解，以及自行加表？

答：1 是的，可以参考给出的表的字段，自己设计额外需要的表去关联

2 每个字段后有说明，您能用上的仅仅是 ID 主键，有些字段应该用不上，所以设计额外的表存储您所需要的字段

1.问题：权限问题

是否可以提供拥有讲师权限的文思海辉教育平台账号以供前端设计参考，从而保证产品兼容性和设计一致性？

答：

账号 fuchuang1 密码 fuchuang1

账号 fuchuang2 密码 fuchuang2

账号 fuchuang3 密码 fuchuang3

账号 fuchuang4 密码 fuchuang4

2.问题：教师视频处理过后的语音是否全部由语音模型生成

如果不采取录音方式，即通过文本转语音来修改视频，是否需要将视频的语音改为全部由语音模型生成，如果不是那么是否需要语音模型的声音与老师的相似？

答：录音方式是为了降低智能语音难度而出的方案，你们可以尝试去实现模拟机器语音与老师语音相似

3.问题：语音合并

您好！我们团队有一个疑问：首先我们在阅读赛题手册时，发现如果要修改语音片段是提供录音功能重新录制，然后将重新录制的语音合并到原视频中，但是在观看赛题视频时，发包方说的是如果想要修改语音片段是先去修改文本，再根据修改后的文本生成语音合并到视频中。这是两种截然不同的方案，请问以哪个为准？

答：这是一个方案，录音功能重新是录制被转换后的文本片段，就是先修改文本，在根据修改后的文本生成语音再合并到视频中，如果难度太大，就靠人工将需要修改的文本片段录音在合并到视频中。

14.问题：录音和文本

命题背景中是将录音合并到视频中，但是任务要求是将文本转为语音再合并到视频中。

答：这是两种技术方向，一种靠人工智能将文本转为语音，如果能通过 AI 做到，就不需要在线录音替换。第二种，如果不能通过 AI 做到，那么再将语音提取文本之后，讲师可以查看哪些文本语句表述有问题，可以针对表述语句进行再次录音修改。应该尝试第一种，将修改的文本转为语音，再合并到对应的视频中。

15.问题：关于系统角色的问题

请问系统是需要一个管理员和普通教师角色吗？还是只需要教师不需要管理员

答：普通教师即可，有测试账号 参见问题 1

16.问题：关于需求问题的提问

1.不添加字幕。需求答疑视频中您说语音转文本不是添加字幕功能，那么如何体现语音转文本功能；

答：不是添加字幕。将视频流中的音频流提取，靠算法实现文本转换，就是语音转文本片段。要求标注起始帧结束帧，时长。

2.录音功能。视频语音提取系统要求将视频中的音频分离出来，通过语音转文本的方式转成文字，但又提供了录音功能，也就是说教师可以通过录音替换指定的音频片段，那么语音转成文本的意义在哪里？

答：这是两种技术方向，一种靠人工智能将文本转为语音，如果能通过 AI 做到，就不需要在线录音替换。第二种，如果不能通过 AI 做到，那么再将语音提取文本之后，讲师可以查看哪些文本语句表述有问题，可以针对表述语句进行再次录音修改。

1、问题：语音修改的方式

在项目文档中，要求有录音功能。而在介绍视频中，说需要文本转语音。所以想请问应该是哪一种？是用户通过录音重新录入呢？还是用户只更改文本，软件自动将更改后的文本转化为语音？

答：两种行为都要支持，录音功能是在你们无法实现 AI 文本转语音的前提下，最后的行为，优先文本转语音，有精力在加上录音也无妨

2、问题：视频的数据接口地址是什么

数据接口的说明是 文思海辉已有的在线课程

是否有类似 API 的数据接口 去直接使用视频资源 还是需要自行下载视频

答：自己找个视频最快，我们的地址 www.ommdata.com，我们的视频不能直接对接你们的平台，你们的平台是完全独立新的系统

【A11】 Web3D 短视频学习教育平台 WSVTP 【文思海辉】

1、问题： 教学课程设计内容是否直接和提供接口数据相关？

答：

部分功能需要用到数据接口

2.问题： 是否有固定指向 譬如 有关软件设计 java php 等教学内容？ 还是自己定任意领域教学内容？

答：

有 JAVA 做服务。达梦数据库存储数据 数值分析和计算。可以用 Python 但最后要转为 JAVA 模型

3.问题： 数据库结构

是否必须将数据库表的所有字段都用起来。还是可以根据自己设计的逻辑，只选择一部分字段？

答：1 根据自己的逻辑，去关联给出表的主键字段，使用哪些字段，您可以自行选择，不必都用，自己设计额外的表

4.问题： 提交材料中 3D 门户以及工作室绘制算法具体是指什么呢
提交材料中 3D 门户以及工作室绘制算法具体是指什么呢？

答：3D 门户是一个基于 Web3D 技术绘制的交互图，请描述您是具体怎么描绘这个 3D 页面的，要详细如：数据如何定义，如何计算的数据，如何将数据应用在 Web3D 技术框架上，目标是完成了哪些绘制和交互任务。在空间时间复杂度上的结论是什么

1、问题： 3D 模型导入 three.js 中遇到问题

1.3dmax 软件导出 obj 文件太大了，三四百 MB，网上有些清晰的模型也才几 MB 或几百 KB，不知道该怎么减小 obj 的大小。

答：是否试过使用 Blender？

文件体积大，延长了加载时间，请在页面上给出预估的加载进度。

参考链接: <http://www.moxingyun.com/news/detail-jichu-1008798371.html>

2.我们现在通过减少 3d 模型的面数来减小 obj 文件大小。但会使导出的 3d 模型的 mtl 文件材质数减少，比如：一个书桌原本有两种材质，桌面+电脑，就可以贴两张图。现在减少面数，导致一个书桌只有一个材质，桌面和电脑一样的贴纸，这就有问题。

答：界面提示加载进度

参考链接: <http://www.moxingyun.com/news/detail-jichu-1008798371.html>

3.所以如果要减少 obj 的大小，那么加载出来就不好看，甚至一大片的纯色图案。如果不减少 obj 大小，加载都会出问题。我们尝试过 draco 文件加载，但是问题通过网上的办法都解决不了，感觉相关的资源很少。于是我们现在准备用 fbx 文件加载，这种文件自带贴图，更加真实，现在遇到的问题就是用 3dmax 导出的 fbx 内部的贴纸一部分丢失了

答：1是否尝试过 Blender 制作 obj 文件？

2 文件体积大，延长了加载时间，请在页面上给出预估的加载进度。

3参考链接: <http://www.moxingyun.com/news/detail-jichu-1008798371.html>

2、问题：3D 门户问题

我看到文件里要求企业提供信虚拟大学以及学院的部分信息。如信创大学等。我想请问大学名字 还有学院是由我们自己决定的吗？

答：我只是给个例子，具体名字你们自己决定哇，参考

我们的平台: <https://www.ommdata.com/>

1、问题：上传核验任务资料

在虚拟屏幕上上传核验任务资料是什么意思，是需要学者完成一个任务的代码上传并检验代码是否正确吗

答：有两种情况，一种较为简单的可以通过程序运行检查结果的核验，如远程编译执行，第二种，目前就是靠后端的人工方式核验，谁发布的任务，谁去检查

2、问题：JAVA 和 three.js 的要求掌握那些技术

需要使用到的 Java 技术以及 three.js 的具体技术的掌握

答：three.js 可以 web3D 虚拟建模，是一个完整的工具包，Java 就是制作后端数据服务接口 springboot+springmvc+jsp+database(rdbms,nosql),尽量后端服务只返回 json,前端 three.js 加其他 Javascript 框架处理交互事件。

【A12】基于手势识别的会议控制系统【长安计算】

1.问题：控制手势和功能应用

1.是否可以自己设计手势实现五种基本的功能。

答：

五种基本动作手势为题目要求必选项，选题者可在这五种基本手势动作的基础上进行扩展。

2.对于五种基本功能是否需要考虑其具体应用场景。例如，抓取功能在日常会商会议中具体会用在什么地方。

答：

需要考虑。

3.问题：演示系统该根据分类结果做出何种显示？

通过模型分类出结果后，控制系统该做出何种反应，例如分类为旋转手势后，是让 PPT 旋转吗？点击的结果又是什么呢？希望五种基本手势在控制系统的展示都获得企业的解答。

答：会议控制系统不局限于 ppt 的操作，也可包含在会议中进行图片操作，即在会议中存在对通过点击等操作进行系统交互，完成打开、选取图片等相应操作界面的确认工作；通过平移、旋转、缩放等完成对图片的一般交互操作，通过抓取实现图片及其他组件的自由移动操作。

4.问题：关于视频中的人数问题

请问贵公司要求完成多人检测功能吗？200ms 的检测时间是只需要对单个人检测就可以了吗？

答：要求实现多人检测，200ms 时间只针对目标演示者的每一个手势动作的检测和识别时间（即从执行完手势动作到输出结果之间的时间）

5.问题：对会议演示系统的实现有要求吗

赛题仅仅是考虑手势识别结果和相应应用场景，还是说要完整实现可交互的会议演示系统？

答：赛题侧重在会商演示背景下的手势识别结果，重点考察手势识别算法的性能，演示系统需要实现，但只作为算法的展示环境，即手势识别的演示场景，可以保证相关动作识别的展示即可。

6.问题：数据输入是 3D 坐标信息吗

最终评分是出题方指定的一些普通 2D 摄像头拍摄的视频，还是手部关节的 3D 坐标信息等自定义的输入数据信息？

答：2D 摄像头拍摄的视频。

7.问题：题目要求中的多人交互问题

手势识别是否要求实现多人识别？如果是，对于多个目标的识别结果，是否需要考虑优先关系？

答：要求考虑多人识别，可考虑场景为每个用户面对一个摄像头进行演示，但因座位关系或场景限制，可能会拍摄到目标用户以外的人。在这个场景下可以摄像头中心的人优先。

8.问题：手势识别考虑身份认证吗？

开会的时候如果摄像头中有多人，则考虑用身份认证来认定只有主持人员的手势可以发挥功能其他人虽然能识别到但不可对 PPT 等进行操作。

答：对身份认证不做硬性要求。可考虑场景为每个用户面对一个摄像头进行演示，但因座位关系或场景限制，可能会拍摄到目标用户以外的人。在这个场景下以摄像头中心的人优先。如果认为身份认证有助于提升识别准确率也是可以的。

9.问题：是否需要写出一个完整的控制系统

您好，请问本题是需要完成一个完整的控制系统？还是仅需要完成一套手势识别算法，或者是需要将该算法集成至现有的会议演示系统中呢？如果是需

要集成至现有的系统中，请问该系统是预设提供的，还是可以自由选择呢？
谢谢！

答：赛题侧重在会商演示背景下的手势识别结果，重点考察手势识别算法的性能，演示系统需要实现，但只作为算法的展示环境，即手势识别的演示场景，可以保证相关动作识别的展示即可。

10.问题：关于会议形式以及操作者所在位置等问题
您好，关于本题目我们有如下几个问题想要请教一下：

(1) 会议的形式是怎么样的,是线上还是线下，线下的话是像传统的前面一个大屏幕,演示者面对大屏幕操作(一个识别终端)还是参加会议的人可以对着自己的电脑摄像头进行操作(多个识别终端)来与其它人共享操作结果？

答：线上线下均可。线下场景指参加会议的人可以对着自己的电脑摄像头进行操作(多个识别终端)，在这种情况下可认为线下与线上场景具有一致性。

(2) 对于摄像头识别到的动作以及人体姿态要求是否可以按照五种手势动作演示视频那样作为最低标准？

(3) 对于硬件设备是否具有要求限制?因为不同的 GPU 对于识别速度是有很大的影响的。

答：识别时间上限为一个手势动作的检测和识别时间（即从执行完手势动作到输出结果之间的时间）不超过 200ms，在此基础上不将识别速度的高低作为评价指标。

11.问题：摄像头

最终实现对手势识别的传感器或摄像头有什么要求吗？如果没有要求，最终评分如何进行？控制系统采用的摄像头是笔记本摄像头还是外置摄像头？如果是外置摄像头的话是否可以使用 FPGA 芯片？所用摄像头可以借助深度摄像头或红外摄像头吗，还是只能用普通摄像头？

答：为公平起见，只使用普通 RGB 摄像头。

12.问题：传统会议场景也是自己设定吗

题目中只给出传统场景中存在的痛点，以及期望等到的效果而没有给出具体场景设定。传统会议场景也是自己基于题设要求设定吗？

答：场景可考虑为每个参会者拥有一个摄像头及相关处理设备用于拍摄、处理各自的手势并在屏幕终端进行显示。

13.问题：题目

请问这个题目是算法类题目，只要设计算法能识别 5 个手势动作就可以了吗？还是在识别手势动作的基础上，要实现人对着屏幕，屏幕会根据动作做出一定的反应？

答：只需考虑给出的 5 个手势动作，但需要考虑避免对非手势动作的误识别。

14.问题：展示形式

展示形式是什么呢？是隔空触控操作投影屏幕吗？

答：场景可考虑为每个参会者拥有一个摄像头及相关处理设备用于拍摄、处理各自的手势并在屏幕终端进行显示。

15.问题：五种基本手势中的抓取到底是指怎样的操作？

在 pc 端，抓取似乎和点击、缩放过于综合，如果不分清楚的话，可能会影响到具体评分，这种情况下该如何抓取和缩放？

答：在动作上，抓取是单手操作，缩放是双手共同操作；在功能上，抓取是抓住某个组件，后续可进行上下左右移动，缩放则是对展示的图片等进行放大和缩小。

16.问题：关于会商演示系统的交互问题

企业方给的【业务背景】提到本题主要“针对基于增强现实多人协同研讨”，请问会商演示系统是基于增强现实的吗以及是否提供会商演示系统？若不提供，需要实现会商演示系统吗？若需要，能说明一下具体需求吗？

答：这里指的是通过手势识别实现非接触式人机交互。赛题侧重在会商演示背景下的手势识别结果，重点考察手势识别算法的性能，演示系统需要实现，但只作为算法的展示环境，即手势识别的演示场景，可以保证相关动作识别的展示即可。

17.问题：基于手势识别的会议控制系统

我们捕捉动作，能不能借用 Kinect 仪器？

答：为公平起见，只使用普通 RGB 摄像头。

18.问题：训练模型的数据集

老师你好，A12 题手势识别中要求提交详细的设计方案，包括模型训练方案，请问有提供用于训练模型的数据集吗？

答：为保证和基于传统计算机视觉的方法的公平基础，不提供训练数据集，可根据样例视频自行拍摄相关数据进行训练。

19.问题：功能问题

点击功能，是否需要明确具体点击哪里吗

答：暂时无需考虑点击位置，只考虑点击动作。

20.问题：功能问题

抓取和旋转具体需要抓什么东西，和旋转是对演示的页面旋转吗

答：会议控制系统不局限于 ppt 的操作，也可包含在会议中进行图片操作，即在会议中存在对通过点击等操作进行系统交互，完成打开、选取图片等相应操作界面的确认工作；通过平移、旋转、缩放等完成对图片的一般交互操作，通过抓取实现图片及其他组件的自由移动操作。

1、问题：数据集

企业能否提供更多数据集以提高系统鲁棒性

答：为保证和基于传统计算机视觉的方法的公平基础，不提供训练数据集。

1、问题：最后提交结果检测方法

请问赛方最后对我们提交系统检测的方式是通过视频集还是实时操作

答：两种方式均会采用。

2、问题：关于各种手势的具体动作限制

请问例如缩放旋转等动作是否一定要参照范例视频中的样式，比如缩放可否通过单手动作实现，这样子更加美观，符合会议场景。

答：必须参照范例视频中的样式。

【A13】聚焦人脸识别的大数据测试系统【虹软】

1.问题: 是否有 baseline?

是否有 baseline?

答:

没有

2.问题: 数据集是否有指定?

数据集是否有指定? 用什么?

答:

数据集没有指定的, 可以从网上下载, 展示时最好能说明一下数据集构成

3.问题: 以服务端形式是在服务器环境上吗?

以服务端形式是在服务器环境上吗?

答:

是的

4.问题: 大数据分析系统

想问下关于大数据分析系统的开发方面, 贵公司有没有建议的开发框架, 小组成员可以借鉴学习

答: 没有特别建议。

5.问题: 资源

请问开发的资源能否用阿里云或者与之类似的资源, 如果不能, 贵公司能否提供一些建议或者说对于资源这一块有什么具体的要求

答: 没有限制。

1、问题: 数据库

数据库可以用 MySQL 么

答：可以。

2、问题：实时处理
需不需要实时处理

答：可根据设计的功能来确定。

3、问题：分布式
该系统的分布式的作用具体是什么

答：有效利用系统资源，提升整体执行效率。

1、问题：人脸识别 SDK
请问性别预测能否设置阈值，如果能，请问如何设置？

答：没有阈值

【A14】 移动互联时代的设备管理【虹软】

1.问题：人脸识别技术的使用
企业提供的资源需要企业认证，请问如何才能通过认证使用？企业会帮助完成认证，进而使得资源可以下载吗？

答：官网有提供免费版本可以个人认证直接使用。

2.问题：设备归还是否需要设计流程
需要设计设备归还的流程吗？设备归还需要经过检验再归还吗

答：可以考虑在内。

3.问题：设备范围
请问管理的设备有哪些类型，针对哪种行业？

答：设备类型不限，如测试仪器等，可主要考虑便携的设备。

4.问题：赛题中系统面向的是什么样的公司，一般是什么设备

赛体中提出的系统是面向什么样公司？公司是从事生产还是其他的，开发手机之类的？需要管理的设备是怎样的，一般这些设备统一保管在一个仓库还是存放在每个部门？设备除了手机具体有哪些？

答：设备类型不限，如测试仪器等，可主要考虑便携的设备。

5.问题：RFID 读写设备是否有指定设备要求

你好，我们团队在进行 RFID 开发时发现，不同的 RFID 读写设备采用不同的通信协议。想咨询一下贵公司，对于 RFID 的读写设备有没有要求，是需要基于哪种协议的设备，还是任意一款设备都可以。

答：没有特定要求。

6.问题：关于盘点

设备盘点是只需要盘点在库的，还是那些被员工借出去的设备也需要盘点

答：可以都考虑在内。

7.问题：设备管理平台是通用还是专用

是要做一个通用平台，考虑到不同公司各种设备，还是针对自己定义的某个行业

答：通用型即可，可主要考虑便携设备。

1.问题：RFID 具体情况

请问 RFID 需要开发在移动设备中还是可以用现有的 RFID 读写设备

答：现有的 RFID 即可，如有更好的思路，也可以体现出来。

1、问题：关于用户期望的问题

在用户期望中有提到“设备非正常带出公司时，自动报警并记录”，请问关于判断设备是否带出公司的依据，是可以在被管理设备上安装 GPS 或其他可能的装置来判断，亦或是团队自由确定的其他方法等。

答：团队可自由确定方法，但希望能够设计低成本方案。

2、问题：关于开发说明的问题

在开发说明中有提到“要能同时管理上万台设备”，请问意思是一共有上万台或更多的设备需要管理，还是系统需要考虑可能存在的高并发操作。

答：管理上万台同时考虑并发。

3、问题：需求具体场景

请问赛题需求是否有具体场景？管理便携设备还是难移动的设备，管理的设备的安全包不包括设备的意外损坏的监控？

答：可主要考虑便携设备。不必要包括设备的意外损坏监控，但需要能记录查询设备的状态。

4、问题：设备短借问题

设备短期借用时，比如五分钟借用，这个问题即使有一个现成的管理系统，也难以实现完美管理，请问这个是否需要考虑？

答：可以主要考虑借用记录的便利性，方便使用的系统使用率才会高。

5、问题：重要设备的定义以及具体例子

请问重要设备是指哪种？重要设备是否指不能带出公司的设备，非重要设备是否指可以带出公司的设备？设备是否便携？能否多举出一些具体设备例子？谢谢！

答：设备重要程度一般是人为定义，可以自行思考，比如特别贵重的设备。

1、问题：无法使用 ARM 版人脸识别

现项目准备使用 arm linux 开发版进行人脸识别，但是贵司的 ARM 人脸识别（对比）需要企业认证才可以进行试用，请问是否可以给一个 arm linux(1109)试用权限或者是否可以使用其它厂商提供的人脸识别（对比）技术。感谢！

答：可以提供 1109 的试用码，可以发邮件到 jjc9523@arcssoft.com，备注大赛名称以及队伍名称和联系方式

【A17】金融领域的事件因果关系抽取【同花顺】

1.问题：关于训练模型和数据获取的问题
请问在训练模型时，使用自己的数据进行训？

答：可以使用自己的数据进行训练

2.问题：后续是否会给出更多已标注好的样例？
对于已给的数据集，我们觉得数据量不够，需要更多标注好的样例和测试用例检验我们的模型，所以请问出题方后续是否会给出更大的数据集，抑或是要求我们自己去找？

答：无法提供更大的标注数据集，可以提供更多的未标注样本

3.问题：我们需要出题方提供详细的标注说明
在我们后续的开发中，肯定是需要检查结果的正确性，但我们无法保证我们理解的正确的标注就是出题方认可的。所以我们需要出题方对于样例的详细标注说明。

例如：

- 当一个句子由多个因果关系时，是对同一句子重复标注还是只标注一个
- 判断一个句子具有因果关系的最低标准是什么，比如如果缺少中心词是否继续标注

答：当一个句子有多个因果关系时，会对同一句子重复标注，但每次标注的中心词不一样
缺少中心词的数据认为无因果关系，无需继续标注，但有中心词的句子不一定包含因果关系

4.问题：json 文件的 text 是否有长度限制？json 文件的 text 是限制一句话之内吗？

答：json 文件的 text 是一句话，长度限制 512

5.问题：能否用自己收集的数据对官方提供的数据集进行扩展？
能否用自己收集的数据对官方提供的数据集进行扩展？

答：可以

6.问题：官方是通过什么软件对数据进行打标注的？
官方是通过什么软件对数据进行打标注的？

答：自己开发的标注工具

7.问题：是否需要前后端？

答：不需要

8.问题：因果关系抽取的问题
分为显式因果关系抽取和隐式因果关系抽取，是否要进行隐式因果关系抽取？

答：无需进行隐式因果关系抽取

9.问题：关于所用测评集内容的问题
最终给出的 2000 条评测集都是跟样例一样，既是金融事件也有因果关系吗？

答：对的，既是金融事件也有因果关系

10.问题：是否有验证集用于验证模型的准确率
赛方你好，请问有没有验证集，给我们验证模型的准确率？

答：无

11.问题：一个句子中是否可能包含多组因果，又是否存在因果嵌套等
注意到带有标注的数据都是一句话一组因果标注，然后人为从句中看的话，
可以清晰看到有的句子可能存在多组因果关系。同时经常存在一组因果中的
果中又包含一组因果；一组因果中因和果可能有多，即所谓原因核心名词、
谓语有多个而只有一个结果，或者反之。这些情况如何处理。

答：测试时，句子会提供中心词，一个中心词一般只对应一组因果关系，无需考虑因果嵌套问题，一组因果中的原因和结果都有可能是一个或多个，需要抽取出该中心词对应的全部原因和全部结果，中心词可以自己挖掘一些包含因果关系的关键词，比如：导致、造成、由于等

12.问题：关于测试模型的方法

请问您会提供测试集吗，还是等提交模型后在您那边做测试？您的测试集会关注到金融领域以外的文本吗，或者说您关注模型在泛化领域上的表现吗？

答：不会提供测试集，需要提交模型代码，然后在后台做测试，测试集只包含金融领域以内的文本。

测试时输入一条 json，同标注格式，其中中心词已标记出来，带有位置信息，输出格式也一致。具体测试代码的标准格式会在之后公布。

13.问题：关于最终模型评价所用到的 2000 条数据的问题，以及有关“中心词”的问题

(1) 请问最终评测模型的 2000 条数据来源在哪里？另外这些数据会提供给参赛选手吗？

答：来源为经过筛选的包含因果关系的金融文本等，不会提供给选手；

(2) 50 条标注样例中出现了“中心词”，请问最终的评测结果跟“中心词”有关吗？

答：测评数据提供中心词，但中心词不参与测评分数计算

14.问题：有关文本预处理与标注

对语料的文本预处理是什么意思，标注一定得通过程序算法吗？可不可以直接人工找出 800 条的因果关系再把它转成 json

答：提供的代码文本预处理需要完成对 json 格式的转换，可以人工找出 800 条因果关系再把它转成 json

15.问题：数据标注相关问题

50 条标注信息中，第 5 条的“技术壁垒等因素”中的“等因素”被标注为原因中的核心词，而第 32 条中“日内择时交易等原因”中的“等原因”没有被标注为原因中的核心词？

答：存在标注标准不够统一问题，后续会重新提供 50 条重新制定标注规范的标注数据

16.问题：数据标注问题

50 条标注信息中第 34 条，为什么不标注为：结果中的核心词：公司，结果中的谓语与状态：加速发展？

答：参考第 15 问

17.问题：数据格式问题

赛题方提供的 50 条标注信息中，start_block、end_block、sub_answer 分别代表什么意思？

答：对赛题无影响，可以忽略

18.问题：数据标注相关问题

50 条标注信息中，第 38 条结果中的核心词标注为公司业绩，而第 43 条结果中的核心词标注为业绩，而不是公司业绩？

答：参考第 15 问

19.问题：关于最后提交的结果的问题

在给的数据中有 5 个论元比题目描述中多了一个中心词，请问最终需要提交的结果是 4 个论元还是 5 个论元，还有最后的提交的结果的格式是什么样子的

答：比赛结果提供代码，测试集输入时已经给定中心词，所以中心词不参与比赛分数计算

20.问题：数据集扩充问题

题目只给了 50 条标注的文本以及 800 条未标注的文本 是否可以自行扩充数据集来保证更好的 f1

答：可以自行扩充数据集，后续会酌情提供更多的未标注样本

21.问题：关于推理速度的问题

评价模型的指标是 f1 和推理速度，而我们想知道推理速度最后怎么进行验证？有没有一个硬性的指标，比如最起码需要到达一个多少多少 fps 这样的指标？

答：推理速度是通过提供代码本地测试所得，没有硬性的指标，但是会作为最终成绩的因素之一

22.问题：数据标注相关问题

预料中存在关系对但没有中心词的情况是否存在？

答：不存在

23.问题：有关给出的 50 个标注标注里面的 key 字段是什么意思

答：是样本的唯一标识符，可忽略；

24 问题：数据相关问题

(1) 指导书描述的只需标注 4 类标签，而给出的标注示例中却有核心词 5 类标签，实际按哪个评分？

答：实际按 4 类标签评分，中心词标签会在测试集输入时提供

(2) 标注样本中，block_id, sub_answer, status 等标签什么意思？能不能发一份具体含义的说明书？

答：block_id, sub_answer, status 为标注软件的标识符，可忽略

(3) 测试集的格式是怎样的？测试时要求输出格式是怎样？评测是如何进行的？上传模型还是别的？

答：测试集的格式和标注样本的格式一致，其中中心词的信息已经标注好，要求输出格式也是标注格式，评测是提交模型代码后后台测评。具体测试代码的标准格式会在之后公布。

(4) 测试时每一条样本都包含因果关系吗？是否有嵌套因果关系？

答：每一条样本都包含因果关系，不含有嵌套因果关系

25.问题：额外数据使用

请问除了发布的 800 条无标注数据集外，可以使用额外的无标注数据集么？

答：可以

26.问题：数据标注相关问题

“表明”是否算中心词

答：由自己判断，中心词的挖掘也作为本赛题的一个考核点

1.问题：关于预测代码的问题以及 CUDA Toolkit 版本问题

1. 请问给出的预测代码中，样例输入带有中心词。那么在最终测试时，测试输入也是会有中心词吗？不需要我们去抽取吗？

答：不需要自己抽取，我们测试数据会在 qas 字段提供中心词

2. CUDA Toolkit 版本都支持哪些？

答：9.0,10.0,10.1

2.问题：数据标注相关问题

新给的 276 条训练样例与之前提供的 50 条样例标注的规律不统一，比如之前没有出现过原因中的核心词为空缺的情况

答：按照新的 276 条标准

3.问题：数据标注相关问题

企业最终用来测试的 2000 条数据，会给出关系的中心词嘛？

答：会的，会在 qas 字段中提供, qas:[[{“question“:”中心词”, “answers“: [xxx]}]]

4.问题：测试相关问题

请问最终给的 2000 条测试集中，每条语料只有一条关系 还是有多条

答：每条语料只会给出一个中心词，一个中心词可能会对应多个原因或者多个结果

5.问题：代码格式

必须实现 predict 函数，输入输出不能修改，是指必须使用 Predictor.py 这个文件里面的代码吗？

答：不一定，但是必须按照规定的 Predictor.py 中 predict 接口限定输入输出的格式去做模型预测

6.问题：数据格式问题

代码文本预处理要完成对 json 格式的转换，那为何要求输出格式又要与标注样例一样？两者格式不是不相同吗？

答：预测时，模型的输入是一条 json 格式的数据，其中 qas 字段会给出中心词，预测结果要求返回一个包含原因、结果以及中心词的 json 格式的数据，至于中间过程的数据处理可以自由发挥。

7.问题：数据标注相关问题

预测代码中，对于输出结果为 True 的情况，是怎样给分的？是否每预测为 100%正确才给分。

答：会统计每一个问句的准确率和召回率，以及统计每一条输入因果关系对的准确率和召回率，并不一定这句话的因果关系完全预测正确才会给分。

8.问题：关于数据集各种名词

数据集中有各种名词，如 block_id, key 等，能提供对这些词的解释吗？

答：block_id 和 key 字段是我们用来定位和标识每一条数据的，打比赛用不到这些字段；

9.问题：数据标注相关问题

在 276 条标注数据中，存在着指代现象，第 56 条、第 159 条和第 198 条中，都进行了指代消解，但是在第 110 条、第 184 条、第 198 条中，核心名词为代词（其）或者为空，这是因为文本中没有提及相应实体对象吗？如果是，那么标注时遇到这种情况，是要填写该代词还是为空呢？

答：遇到指代的情况，标注为空，不需要标注代词，后续我们会对测试集这种情况进行清洗。

10.问题：数据标注相关问题

在 276 条标注数据中，第 147 条“自住型、改善型需求”与第 233 条“重视销售、依赖拉新的模式”结构类似，但是一个是没有分开的，一个是分开当做两个核心名词的，这是为什么？

答：标注错误，应当放到一起标注，因为名词为需求或模式只有一个，前面的字只是修饰这个名词，当出现两个名词时，应当分开标注，如重视销售的模式、依赖拉新的模式，这种需要分开标注。

11.问题：数据标注相关问题

在 276 条标注数据中，第 102 条中“一些地方逢喜事就办酒宴之风”把“一些”归入核心名词，第 234 条中“：一些企业”没有把“一些”归入，标注时遇到这种数字、量词该如何处理？

答：一般数字和量词是不纳入标注范围的

12.问题：测试集中中心词数量限制及多对因果关系对如何标注。

测试集中一句话对应中心词的数量最多为多少？标注数据时，如果有多对因果关系对，是否均要标出

答：测试集每句话只有一个中心词，如果一个中心词有多个关系对，需要将它们标出

13.问题：一些特殊情况的如何标注

像 A 导致 B，从而导致 C，其中 B 即是结果也是原因，该如何标注；像 A 导致 B，并且 A 影响 C,这种原因相同，中心词和结果不同如何标注

答：这种形式在测试集没有，不用考虑

14.问题：标点，‘等’是否标注

像 A、B、C 等原因导致了 D，是否标注“等原因”，以及标注的部分中是否可以出现标点符号。

答：不需标注“等原因”，标注时可以出现标点符号，但当 A、B、C 属于三个不同的名词时，则需要分开标注，中间顿号不需要标注。

15.问题：是否关注一句话中的非金融事件

如果一句话中有非金融事件因果关系，是否需要标出

答：需要标出，模型关注金融文本中的所有因果关系，不一定是纯金融事件的因果关系

16.问题：如果“原因（结果）的核心名词”和“原因（结果）中的谓语或状态”分别有多个，是否需要顺序对应

答：当前模型不需要做顺序对应，需要给出每一个问句答案在原文中的 start 和 end 位置

17.问题：何时给出较为详细的标注规范
何时给出较为详细的标注规范

答：对于金融领域的数据，很难去找到一个标准去定义所有的句子语法类型，所以标注标准只能是从已有的数据中寻找一定的规律，但是如果你们发现标注中的问题，我们会解答疑问并根据解答对测试集做出一定的清洗工作。

18.问题：数据标注相关问题

在 276 条标注数据中，第 51 条“建设大连金普新区”与第 187 条“打消就业担忧”结构相近，但是 51 中将之分离，“建设”谓语，“大连金普新区”作为名词，而 187 条中这一段话当做谓语，这个标注的规则是什么呢？

答：应该分开，就业担忧作为名词，打消作为谓语，后续我们会对测试数据进行清洗。

1、问题：评测标准以及标注问题

1 请问评测时的对数是指哪个维度呢，是 token 级别的评测、短语级别还是对数呢？

答：关系对，是指（原因名词，原因谓语，结果名词，结果谓语）组成的元组。

2 请问当一句话中包含两组因果关系时，是否都需要标注？

答：可以当作两个数据标注，每一个数据标出一组因果关系，两个数据的因果关系中心词不相同。

2、问题：标注过程中是否要标出核心名词和谓语或状态的详细信息

平均银行及其他借款及融资租赁安排的租赁负债融资租赁应付款项总额相较 2018 年年度上升导致财务费用增加；

这句话中的‘原因中的谓语或状态’是要求标出更详细的信息，如“相较 2018 年年度上升”，还是要求只标出“上升”即可？

答：标出上升即可，2018 年年度是一个时间状语，一般不需要。

全球“火箭式”流动性释放导致名义利率持续下行。
这句话中的‘原因的核心名词’是要求标出更详细的信息，如“全球“火箭式”流动性”，还是要求只标出“流动性”即可？又如是需要标出“汉高集团销售额”，还是只需标出“销售额”即可？

答：全球“火箭式”流动性和汉高集团销售额。

3、问题：有关标注格式

可以将赛方给的标注格式 A，通过代码改成另外一种标注格式 B 以便于训练，最后再将输出弄回格式 A 吗？这样算不算离题

答：可以，只要输入输出的格式按照规定的格式就行，中间的数据转换过程可以自由发挥，提交的代码中需包含中间的数据转换过程的代码。

4、问题：中心词标注

在上次答疑中说”当一个句子有多个因果关系时，会对同一句子重复标注“，意思是重新标注多个 json 数据，还是在一个 json 数据中包含多个中心词，意思时一个 json 数据中会有几个中心词？

答：两种方法标注方法都可以，但是在送入模型时，若一条 json 数据有多个中心词，应当拆分成多个数据训练和预测，但是测试集我们会提前处理好这样的数据，保证每一条 json 数据只有一个中心词。

1、问题：对于评价指标的疑问

假设测试集包含 200 条数据，其中 100 条带因果，100 条无因果关系。模型预测正确 150 条数据，其中 70 条带因果，80 条不带因果。那么 $precision=150/200$ ， $recall=70/100$ ，请问这样理解是否正确？

答：不是这么计算的，一条数据会出现原因或者谓语为多个答案的情况，会将它们组合起来，如真实标注“原因名词 1”，“原因名词 2”，“原因谓语”，“结果名词 1”，“结果名词 2”，“结果谓语”，则组合起来会有四种情况：

1. “原因名词 1, 原因谓语, 结果名词 1, 结果谓语” ,
2. “原因名词 1, 原因谓语, 结果名词 2, 结果谓语” ,
3. “原因名词 2, 原因谓语, 结果名词 1, 结果谓语” ,
4. “原因名词 2, 原因谓语, 结果名词 2, 结果谓语”

预测标签为“原因名词 1, 原因谓语, 结果名词 1, 结果名词 2, 结果谓语” , 则组合起来有两种情况:

1. “原因名词 1, 原因谓语, 结果名词 1, 结果谓语”
2. “原因名词 1, 原因谓语, 结果名词 2, 结果谓语”

正确答案为 4 个, 预测了两个答案, 其中两个预测正确:

Precision 为: $2/2=100\%$, Recall 为 $2/4=50\%$ 。

2、问题: 最后验收的测试集和给的训练集一样都是一定包含因果关系的短句吗?

还是说是给大段的文章; 或者在测试集的句子有无因果的干扰句子

答: 测试集中都是包含因果关系的短句, 但是某些短句可能只有名词没有谓语。

3、问题: 数据标注相关问题

在 50 条中, 文本 27 “被撞断掉落”, “被”算入原因中的谓语, 而 276 条中, 文本 91 “被法院裁定受理破产重整申请进入重整程序”, “被”没有算入, 那么标注时遇到这种情况, 要不要加入“被”?

答: 标注时要加入“被”

4、问题: 数据标注相关问题

关于上次答疑情况中, 遇到指代的情况, 如果这个指代在本文中有相应的实体对象, 那么是否要标出? 还是说也是空白?

答: 空白, 这次比赛内容不涉及指代消解

5、问题: 数据标注相关问题

对于量词和数字, 标注的规则还是有些不确定, 比如说是“再次”“再一次”“部分”“整个”这样的词, 那么是否要算入核心名词或者谓语状态呢?

答: 不算入核心名词和谓语

6、问题: 评价指标

主办方，您好。请问能否对评价指标做出事例说明呢？

答：参见问题 1 回答。

7、问题：数据抽取规则

1.比赛最终的标准是抽取出来的因果对，那请问抽取出来的内容是短语还是一个词呢？在您公司所给的已标记的数据中，举个例子：近 800 个人，您有些句子抽取出来是：近 800 个人，而有些句子抽取出来是：人。请问最终的抽取结果到底以单个名词为主还是短语为主呢。

答：短语为主，短语名词一般包含修饰该名词的定语，不包含该名词的状语和量词，近 800 个人应该抽出人。

2.在一些句子中，可能包含多对因果关系，那么请问这种情况是抽取句子靠前的因果对还是靠后的因果对，或者是两个因果对都抽取出来。在您给出的数据中，抽前和抽后的情况都有，所以我们有一些不清楚规则，来打扰您问一下。

答：靠近中心词的因果对。

3.部分语句含有“顿号”、“和”、“及”、“或”等词，是抽取一个还是多个抽取，例如：从市场热点看，温总理在政府工作报告中调低 2012 年经济社会发展预期目标为 GDP 增长 7.5%，导致蓝筹股走势出现分化，银行、钢铁、电力及公用事业等大盘蓝筹股出现在跌幅前列。结果核心名词是抽取“银行”、“钢铁”、“电力”、“公用事业”还是：“银行、钢铁、电力及公用事业”

答：当顿号、和、及前后衔接的词是独立的个体时，如“银行、钢铁和电力”，则分开标注，标注为“银行”，“钢铁”，“电力”，若衔接的词有共同的前缀或者后缀时，如“钢铁、煤矿和电力公司”，则当作一个整体标注。

8、问题：多因果关系中心词问题

针对官方答疑 3 里的问题 3 的回答：当一个句子有多个因果关系时，会对同一句子重复标注，但每次标注的中心词不一样，此处的不一样是指位置不一样的都算，还是指不同的字面意义上的词，比如句子中有两个“导致”，在不同的位置，是算一个中心词还是算两个？

答：位置不一样，两个导致就标成两个数据，分别对每个导致的原因和结果标注，一个导致一个造成也是标成两个数据。一条数据的中心词是导致，一条数据的中心词是造成。

9、问题：多因果关系标注问题

官方第二次答疑的问题 3 回答说：当一个句子有多个因果关系时，会对同一句子重复标注，但每次标注的中心词不一样；而第三次答疑问题 4 回答说：每条语料只会给出一个中心词；两者是否有冲突？如果每条语料只给出一个中心词的话，就无需对句子进行重复标注了

答：不冲突，意思是如果一个句子有两个中心词，当作两条不同的数据标注，参考问题 8 的答疑。这两个不同的中心词应该当作两个不同的样本送入模型学习，预测时也是一样，但这两个不同的中心词还是在一条句子里面，是一条句子的两个因果关系对，只不过中心词不一样，所以模型要学习的地方也不一样。

【A18】债券图表数据 ocr 检测与文本识别【同花顺】

1.问题：A18 债券 OCR 文本识别的任务二有提供数据吗

答：有提供样例数据，在 ocr 测试集中的识别测试集中

2.问题：模型必须自己训练吗

最终提交成果所用模型必须是自己亲手训练的吗。

答：是的

3.问题：表格的*号需不需要检测

在进行文本框坐标检测时*号是否需要检测

答：“*”需要预测，表示为空的“-”不用预测

4.问题：债券标准

债券有没有标准的统一的。还是任意债券表格都可以

答：都可以

5.问题：文本框坐标标准

文本框坐标有没有具体标准，是画出来就可以，还是有一定范围。

答：不超出图片宽高就行，具体的标注标准参赛者可以可视化我们提供的样例数据查看我们是怎么标注的

6.问题：精度要求

精度要求到小数点后几位

答：小数点后三位

7.问题：任务一与任务二之间是否有关联

想问下，任务一跟任务二的完成是否分开？任务二需不需要将任务一检测出来的文本框分割再去识别？

答：分开，单独训练两个模型

8.问题：所给测试数据有问题

在赛题中，贵公司提到对于任务一即文本框检测任务，会有 50 张图片作为测试数据；对于任务二，即文本识别，会有两万张图片作为测试数据。但是贵公司实际给出的数据却明显少于以上数量，请问是后期给，还是我们的测试数据只有这么多？

答：没有提供赛题中的训练数据，只提供了部分样例数据，参赛者可以自行收集及标注模型训练所需的训练数据；后期提交模型评估时我们会单独提供测试集

9.问题：所给测试数据标签有问题

贵公司提供的测试数据中，部分测试数据存在文本框标签不齐全的情况，即我们预测出来的文本框的数量要大于所给标签的数量，这可能会影响最终的准确率，不知贵公司如何应对？

答：这个并非标签有问题，而是标注的问题；比如一行文本有两句话，按空格隔开，我们在标注时基本上按一个框去标注这行文本；但是参赛者的模型可能会预测出来两个框，这

个所有 ocr 任务的普遍现象，没有关系；最终评估我们是以预测框的坐标和测试集标注框坐标算 IOU 计算准确率为准；所以要求参赛者需要可视化我们提供的样例数据，了解我们在标注这些数据时的标准，做到自行标注数据时标准和我们的测试集标准尽量统一

10.问题：关于 ocr 检测的角度
是否需要考虑文本角度的问题，测试集数据是否会出现有倾斜角度的文本？

答：会有倾斜的文本，但是参赛者提交结果时只需提供框的四个点坐标，不需要提供角度结果，不要求一定是矩形框；我们在评估模型时只会按框坐标去计算 IOU 并算出准确率

11.问题：关于开源框架
程序是否可以在开源框架的基础上进行改进？

答：不限开发工具，但是最终提交结果时需提供参考的框架名称、改进方案、参数优化的过程

1、任务一的评判标准是什么？单纯生成坐标这个要怎么评判，怎么验证正确性？图表中框输出是否有顺序？

答：评判标准：召回和准确率；按照预测框坐标和标注坐标计算 IOU (0.5)，算出召回和准确率；框输出是四个坐标点 (左上、右上、右下、左下) 顺序

2、自己去寻找数据集时，数据集有没有特定的范围？金融方向的就可以吗？还是必须得是债券相关？

答：金融相关，最好是债券类型数据，需要大量的表格数据

3、问题：关于最后提交的问题

请问最后是贵公司给出测试集，我们在本地用模型对测试集进行处理后提交预测结果；还是我们按照贵公司的要求将代码封装，提交模型和代码，由贵公司统一测试？

答：提交预测结果即可；但是最终比赛结束，为了验证获奖的同学结果的真实性，需要提交模型和代码，我们统一预测给出最终评比结果

4、问题：演示模型的形式

您好，请问文字检测模型和文字识别模型需要分别部署到前后端框架中进行分别演示吗？还是可以把检测和识别模型结合之后部署到一个前后端框架中进行演示？

答：都可以

5、问题：深度学习框架

请问对深度学习框架有要求吗

答：没有

6、问题：模型数量问题

我们可以通过一个模型去完成这两个任务，还有必要去再写一个模型吗

答：可以只提交一个端到端的模型

7、问题：关于测试集中文字语言的问题

请问出题方最终提供的债券图表所涉及到的文字除了中文，阿拉伯数字，英文以及相应的标点符号以外还有其他语言的文字吗？

答：没有

8、问题：任务二结果提交的问题

任务二要求提交的 txt 标注文件中 img_path_1、img_path_2 之类的是绝对路径还是相对路径？

答：绝对路径

9、问题：关于准确率的评判标准

图片中的标点符号是否准确识别会作为评价 OCR 精确度的标准吗

答：会，但是我们标注时通统一转为英文的标点

1、问题：关于 OCR 项目识别速度

请问保证识别准确率的基础上，识别速度的快慢会影响评分吗

答：不会

2、问题：测试数据

测试数据是否有手写体，测试数据是否有印章

答：没有手写字体，有印章

3、问题：任务一结果提交的问题

任务一提交结果时图像的注释存储在具有相同文件名的 txt 文本文件中，文本文件中每行要包括检测框的四个顶点坐标，想问一下坐标系的原点在哪里？X 轴和 Y 轴是以像素为单位的吗？

答：原点在图片的左上角，以像素为单位

4、问题：关于测试数据的角度问题

初赛以后是否还有二次评定测试集？如果有，文本角度是否均为正常角度呢

答：有评定测试集；文本角度为正常角度（即便原始图片有偏移角度，偏移角度也相当小，本次参赛测试集不涉及旋转 90,180, 270

1、问题：关于数据

企业提供的检测数据坐标是否是真实的坐标呢？

答：真实坐标

2、问题：关于文本框坐标输出顺序

题目要求每个文本框的坐标按照顺时针输出，那么输出文本框的顺序有什么要求吗

答：文本框顺序没有要求

3、问题：实际检测结果、准确率与召回率对最后评分的影响

请问最后在给作品打分的时候，准确率和召回率的高低对分数的影响很大吗，还是说实际的检测结果更为重要？因为我们在训练过程中发现召回率有些偏低

答：检测框模型任务采用准确率和召回率两项指标做最后评比（参照 SQUAD 评比标准）
识别任务采用完全匹配准确率指标做最后评比

4、问题：关于识别的内容

债券文档中盖章的文字是否需要识别？

答：需要

5、问题：TXT 文本命名和内容排序规则

结果需要生成的 TXT 文本命名及其文本内容排列，可以参照 ocr 测试集里面的命名规则吗

答：参照测试集样例提供的说明进行命名

【A19】智能化共享租赁平台【中信银行信用卡中心】

1.问题：关于淘宝速卖通平台接口接入使用的问题

以下出自速卖通品平台的接入要求的文档：

速卖通自研型开发者要求：

必须为企业，需上传开发者企业营销执照。仅限于具有开发能力的速卖通卖家授权接入仅适合于自己企业的店铺（以店铺企业认证信息为准）申请商家接入必须入驻聚石塔

申请入驻需要在产品说明书中提供自研发系统详细功能模块信息（调用 API 信息，系统架构，功能截图，自研发软件著作权，自研发源代码）

请问我们参赛者如何处理？

答：

速卖通主营业务为跨境电商，对接相对复杂，题目中列出主要是做为参考，本次选题的主要创新环节不一定非要体现在支付环节，而且参加竞赛也并不需要完全复制企业实现业务全流程，所以建议参赛同学，可以尽量简化支付环节，网上可以找到一些供个人账号对接的支付 api，或者采用充值虚拟货币到个人账户进行管理的方式实现模拟扣费，着重体现对整个租赁流程的设计实现

2.问题：关于支付，物流，用户信用的第三方接口问题

命题文档中数据接口部分提供的网址需要企业账户才能获取对应的接口服务，这个企业账户贵公司是否提供？或者能否提供更详细的第三方支付，物流的 API 接口文档呢？项目考虑增加用户信用评估功能，请问中信银行提不提供查询用户信用信息的第三方接口 API？

答：

目前无法对外提供企业账户，如实现困难，建议同学们到网上自行查找相关服务接口，支付宝，以及各个物流公司都能找到详细对接 api，如果因为资质问题，无法接入，建议按

实际 api 格式进行模拟实现。由于客户信息比较敏感，有强监管，无法对外提供相应的 api。

3.问题：是否提供金额在银行中转的 API

我们考虑将收取的定金等金额存入平台绑定的中信银行卡中暂存，请问是否提供一个平台的银行卡，以及存入和取出金额的 API？

答：

目前无法提供相应的 api，建议同学自行调研相应第三方支付平台的实现。

4.问题：商品要与中信银行有关吗？

您好，这里的商品指的是类似天猫那样的商品，还是指定是中信银行相关的商品？

答：

商品可以是任意具备租赁属性的物品，同学们可以自行设计。

1.问题：共享租赁平台

订单的全流程管理 需要包括了物流在内的管理嘛 配送需要真实存在吗？

答：物流信息一般由第三方物流公司提供接口查询，技术上实现不复杂，如果因资质流程等问题接入有困难，也可模拟物流信息

2.问题：关于租赁业务中商品种类的问题

我想问一下关于这个平台租赁业务可否专注于某一类商品的租赁业务，比如只做电子产品的租赁；

答：租赁产品不做限制

3.问题：请问您只是想是一个信用卡的租赁系统：还是一个基于信用卡而各种租赁信息都可以发布的平台；还是基于信用卡只由单一的租赁的信息可以发布，但是不同的类型租赁商家都可租赁商家都可以使用这个平台，就类似于微信公众号（有不同的公众号），但是都是在同一个后台管理的）？

答：自营或者针对第三方商户这方面不做限制，根据自己的设计和能力范围完成即可

4.问题：请问我们可以知道贵公司的信用卡业务流程大致分为几个阶段吗

答：卡中心的租赁业务具体实现从逻辑比较负载，但大体上上包含了租前的商品维护，租赁过程中的订单维护（如租金按月扣取、租金扣取不到的一些列业务逻辑处理），租赁订单自然结束或者认为结束的退租流程处理

5.问题：关于项目形式的问题

你好请问该租赁平台，是否限定平台的形式呢？可否以 App 形式呢？

答：没有问题

6.问题：共享租赁业务模式

只局限于 BtoC 模式还是可以 CtoC,BtoB 等模式

答：都可以

7.问题：平台提供方管理员相当于中介吗

平台提供方管理员是相当于淘宝平台会对公司商家收取门面费吗，相当于中介会赚取差价吗

答：toB or toc 可以自行设计实现、不做限制

8.问题：租赁业务实现方式

共享租赁业务是现场使用还是通过运送物流来使用

答：可以自行选择、业务逻辑及功能实现合理即可

9.问题：平台支付

是否可以调用支付宝芝麻信用 授权商户了解信用评估结果以及使用支付宝支付

答：不限制支付方式

10.问题：租赁模式

请问以租代售 共享租赁是什么样的模式呢

答：以租代售一般针对长期使用的商品，即在租赁过程中用户如想永久拥有商品则提供支付渠道让用户补足剩余差价实现购买，共享租赁则是只针对租赁时间段收取租金快速借还，目前市场上也有不少直接将两种租赁方式合并为一种业务流程的应用，如一些共享充电宝的提供商就有这种业务

11.问题：关于实现 B 端厂商商品及订单管理说法不明确的问题

主办方您好，想问一下，针对这个“实现 B 端厂商商品及订单管理”具体是指我们平台开发的这一端向 B 端厂商买货，还是说 B 端厂商直接把货提供给我们（通过我们的平台把商品提供给用户，相当于我们是中介）

答：答：这个问题其实是平台实现 toB 还是 toC 的问题，这方面不做限制，自由发挥

12.问题：有关赛题中以租代售、先租后买两种模式区别的疑问

您好，题目任务要求第 4 点中有如下要求：“平台要满足不同租赁模式的商品配置及订单管理全流程，如先租后买、以租代售、共享租赁。”目前的理解是先租后买是指租户在租用一段时间对物品较为满意，后续可以选择将物品买下，价格中按一定比例减去之前付过的租金；而以租代售是商家通过让用户先有偿试用（即租用物品），之后再购买物品。似乎这两种模式实质上是同一种模式，请问这里的先租后买、以租代售两种模式具体有什么区别呢？

答：业务逻辑上先租后买和已租代售一个强调租一个强调买，其实业务逻辑上实现相似，共享租赁更贴近于真正的租赁业务，市场上也早已有融合业务，如共享充电宝，可充分发挥自己的创造力，在业务上进行创新。

13.问题：关于租赁模式的问题

可以举一下关于先租后买，以租代售，共享租赁的例子吗或者详细描述一下

答：有相似问题

14.问题：平台收费问题

平台是否会从厂商营业额中收取费用

答：一般是从租金中抽取佣金

15.问题：我们扮演的角色???

这个项目具体是实现什么功能 是不是我们这个项目是实现一个由厂商和客户进行租赁购买的一个平台（包括厂商、客户、平台管理员三方面的角色），这个平台是不是在双方完成交易后收取服务费 还是这个平台也充当厂商的角色

答：平台可以是 toC 也可以是 toB，两种方式具体需要承担哪些功能可自行到网上查找资料

16.问题：关于系统的边界问题

您好，我们需要明确整个项目的边界是否有给出？如 C 端客户及 B 端厂商使用时如何进入我们的租赁平台，需要预留接口还是说我们提供一个完整的注册，资质审核等流程？希望能够得到您的答复，非常感谢！

答：具体边界可自行设计，不是核心的功能，如注册登录等可虚拟完成。

17.问题：关于项目中支付功能问题

您好，支付环节我们需要使用什么样的第三方支付平台接口能否给出具体说明？希望能够得到您的答复，非常感谢！

答：支付方式需自行调研，因为目前主流的第三方支付平台都有门槛，且标准不一，如有困难，建议用虚拟账户模拟支付。

18.问题：买家的日常业务操作是通过 APP 还是一定要求是网页

您好，我想问一下买家的日常业务操作，比如浏览商品列表等，是通过在 APP 的形式展示还是通过网页的形式展示。

答：对实现方式不做具体要求

19.问题：关于智能化共享租赁平台项目能否使用开源代码

您好，我看到在项目介绍中并没有关于开源代码能否使用的说明，希望能够解答一下！

答：偏业务应用，没有统一的开源代码

20.问题：共享定义不是 BtoC

完成共享租赁业务是否一定是 BtoC,通过查找资料共享的定义是 CtoC，用户也可以当做商家，BtoC 租赁是只有企业作为商家，所以在确定用户使用范围的话是什么范围呢

答：共享是一个概念，至于具在何种平台实现不做限制

21.问题：能否以小程序的形式实现用户端？

您好，请问可以将用户端做成微信小程序吗？还是一定要做成 app 呢？

答：可以

22.问题：关于具体所需解决的问题

请问出题方，我们所开发的到底是一个全新的共享系统（从产品到平台这个整体），还是仅仅只是需要创造一个平台（共享产品不需要我们决定）

答：我们需要完成的是平台系统的建设，如果能包含产品管理的部分当然最好，如因时间或资源有限，可以着重在可以体现创新业务的环节多做工作，其他功能自行决定是否用模拟演示的方式实现。

23.问题：微服务架构体系的问题

必须用微服务架构体系的框架来实现吗，springcloud 和 jboot 等等

答：微服务架构考察的是对整个系统功能规划是否清晰，模块划分是否合理，也是目互联网企业实现业务的主要技术架构，建议使用，不做强制

24.问题：作品的形式

请问贵公司的出题本意是单纯做一个租赁平台（类似于淘宝，但是主要业务是租赁，以邮寄的方式完成租赁和退还），通过智能化的租赁流程体现智能化。还是租赁场景以面对面为主（类似于充电宝、共享单车）呢？

答：具体的业务实现方式不做限制，可以自由创新

25.问题：关于题目描述中“BtoC 租赁平台”的问题

主办方你好，我想问一下这个平台设计上可以设计为 BtoBtoC 吗

答：有能力设计实现当然可以

26.问题：租赁商品的个数

建立的共享租赁平台是只需要想膜拜那种共享租赁一种商品，还是要是一个像淘宝形式一样的通用多种商品租赁平台

答：自行设计决定，不做显示

27.问题：租赁方式

是否可以像淘宝一样的线上租赁平台的方式，还是一定要像租充电宝的线下有共享点的租赁方式

答：本题初衷是基于电商平台实现共享租赁业务，也就是上面说的线上租赁平台，但如果系统能解决线下共享租模式并有创新亮点也可以

28.问题：客户类型

请问是类似于淘宝那种世界各地租赁物品、物流快递送达？

还是类似于美团那种近距离租赁，？

或者是像共享单车那样到地取货？

答：如 27 题

29.问题：关于共享经济和租赁经济的概念

本赛题以共享经济为背景，但实际要做的是租赁经济。从估值上看，共享经济（C2C）模式的体量是租赁经济（B2C）无法比拟的。而目前的租赁经济如共享单车、共享充电宝其实都打着“共享经济”的幌子，实际上并不是真正意义上“闲置资源”的共享。我想问一下，本赛题的平台一定是 B2C 吗？是否可以 C2C？

答：可以

1、问题：关于运行机制

请问我们需要构建管理层的网页吗，如果需要我们具体要实现那些功能呢

答：后台管理系统最好还是要实现的，基本的商品管理、订单管理还是要有的

2、问题：赛题需要实现 app 端 or 小程序端嘛
赛题需要实现 app 端 or 小程序端嘛

答：pc 端手机端有一边可以实现就可

【A20】基于 FPGA 的车牌识别应用系统【英特尔 FPGA】

1.问题：如何获取公司给提供的资源呢？

您好，我们现在刚开始准备作品，不知道参赛主办方或者是公司那边是否会提供资源？是否会提供开发平台，怎么样获取？非常感谢您的解答，谢谢你

答：您好，【A20】赛项会给各队选手以团队为单位提供 AWCloud 人工智能边缘实验平台相关使用账号、使用手册、车牌数据集等资源，需要各位移步下列网址填写申请信息，我们收到信息后会以邮件形式发放相关资源，获取地址：<https://www.fpga-china.com/fwwb>

2.问题：最终测试集与训练集是同样形式的吗？

最终的测试集图片里也是只有车牌吗？还是类似于马路上摄像头拍的照片，需要找到车牌再识别呢？

答：您好，最终测试集与训练集采用同样形式，都是通过车牌生成器制作的车牌图片。本次比赛重点考察识别准确度及识别帧率，无需从图像中提取车牌再做识别。

3.问题：最终可执行程序的识别图片是从数据集 B 中读文件吗？是打印程序执行时间还是单张图片的识别时间

1、最终可执行程序的输入是数据集 B 文件夹吗？

2、如果是，题设中说使用 Linux time 命令进行计时，求得运行时间，这里的运行时间是指整个程序的运行时间（也就是识别所有数据集 B 中的图片的时间），还是需要打印出每次识别一张车牌图篇的时间。

答：您好，最终可执行程序识别图片是从数据集 B 中读取文件，可以是数据集 B 的文件夹。运行时间是指整个程序运行的时间（假设程序识别速度较快，可在 1 分钟之内识别 60 张以上，则只需最终时间即可。如果识别较慢，比如 1 分钟识别小于 10 张，则建议除了最终时间外，也在每张图片识别完成后做一次时间输出，以便结果统计）。

4.问题：关于最终测试的数据

测试集的数据图片是一种什么样子？是只包含单纯车牌(和训练集类似)，还是车的前部或尾部图片，需要我们先提取车牌位置再识别。车牌数据是否有遮挡，扭曲等干扰？

答：您好，测试集的数据图片与训练集相同，参考问题 2，只是单纯车牌。

1、问题：请问车牌里会有大写字母 I 和 O 吗？

机动车号牌标准《GA36-2007》5.9.1 的规定“序号的每一位可单独使用英文字母，26 个字母中的 O 和 I 不能使用，不允许字母“O”和“I”出现在汽车牌照里，平时汽车上看到的 I 是数字 1 而不是英文 I。

所以请教一下竞赛涉及到的车牌包含大写字母 I 和 O 吗？谢谢！

答：您好，感谢您的反馈，数据集 A 中确实存在不规范车牌号。在竞赛时使用的测试数据集 B 会依据《GA36-2007》规定执行，不使用字母“O”和“I”。

2、问题：关于比赛时使用的数据集

比赛时使用的数据集是否也是像提供的数据集一样由同样的车牌生成器生成，且全为标准的蓝底白字车牌，没有新能源车牌以及其它特殊车牌。

答：您好，最终测试集与训练集采用同样形式，都是通过车牌生成器制作的车牌图片，且全为标准的蓝底白字车牌，无新能源车牌以及其它特殊车牌。

3、问题：车牌数据集中是否包含 O 和 I 两个字母与其他中国车牌不使用的汉字

1、中国车牌中不包含 O 和 I 两个字母，但是提供的数据集却有“桂 I025P1”这张车牌。

2、中国车牌中没有“台”字头与“滇”字头车牌，但是数据集中有“滇 E1A956”和“台 C68295”。

所以询问一下是否赛题中的车牌是否存在 O 和 I 这两个字母；并且是否有中国车牌中不使用的汉字，如“台”、“滇”。如果有的话，请问相比中国车牌多包含哪些汉字。

答：您好！①在竞赛时使用的测试数据集 B 会依据《GA36-2007》规定执行，不使用字母“O”和“I”。②标准车牌如下图列表简称为准，在此基础上增加“台”。③原数据集中“滇”替换为“云”不再使用。

序号	地区名称	简称	序号	地区名称	简称
1	北京市	京	17	湖北省	鄂
2	天津市	津	18	湖南省	湘
3	河北省	冀	19	广东省	粤
4	山西省	晋	20	广西壮族自治区	桂
5	内蒙古自治区	蒙	21	海南省	琼
6	辽宁省	辽	22	重庆市	渝
7	吉林省	吉	23	四川省	川
8	黑龙江省	黑	24	贵州省	贵
9	上海市	沪	25	云南省	云
10	江苏省	苏	26	西藏自治区	藏
11	浙江省	浙	27	陕西省	陕
12	安徽省	皖	28	甘肃省	甘
13	福建省	闽	29	青海省	青
14	江西省	赣	30	宁夏回族自治区	宁
15	山东省	鲁	31	新疆维吾尔自治区	新
16	河南省	豫			

1、问题：请问测试集的图片有多种类型吗？

测试集的图片是都是蓝底白字的还是有多种类型，比如车牌位数不一样，颜色不同？

答：您好！测试集的图片也为车牌生成器生成，都是蓝底白字，位数与数据集 A 相同。示例图片及详细信息如下图所示：



2、问题：远程控制出现问题

在远程操作时，如果出现无法连接上实验主机或者开发板，可以联系谁解决问题？（例如我曾无意间关闭了 NoMachine 的服务，导致 Nomachine 无法显示实验主机界面，这样对于远程用户来说是没法解决的）

答：您好！云主机或开发板相关问题可以邮箱联系：liuke@awcloud.com

3、问题：资源发放问题

资源发放获取方式及资源类型？

答：资源包目前已经发放给所有团队队长的邮箱，如未找到相关邮件，可以联系 liuke@awcloud.com。提供资源内容包括“边缘实验平台实验指南”、“车牌数据集”、“车牌生成器”、“参考工程文件”及平台试用账号信息。

注意：参考工程文件仅供参考，工程中处理车牌并非车牌生成器产生，需自行设计代码。

4、问题：远程使用

比赛结束后远程平台是否还会继续开放？

答：比赛结束后会逐步回收平台资源，敬请谅解！

5、问题：开发工具是否要求必须使用指定云平台？

赛题说明中“6. 任务要求-->【开发工具与数据接口】”中开发工具可以使用 Cyclone V 或其他，并且推荐使用开发工具为“AWCloud C5 FPGA kit 或 AWCloud 人工智能边缘实验平台”。如果参赛团队自己有其他系列的 Intel FPGA（如 Arria 10 或者 Stratix 10）是否可以使用？以及最终测试结果是否以云平台测试结果为准？

答：推荐使用给大家发放的云端资源，如果赛队自备开发板最好为 Cyclone V 系列，如其他超出性能太多的，比如 Arria 10 会对其他赛队不公平，不推荐使用。

最终测试结果，如使用云端资源的，则用云主机直接测试，其他情况则需现场测试。

6、问题：FPGA

openc1 实现神经网络放大后然后如何部署到 FPGA 上

答：openc1 在我们的云端资源上是不支持的。

7、问题：指导书

指导书内的内容不知道怎么做，找不到指导书里所说的文件

答：这个问题应该是指《人工智能边缘实验平台-实验指南(200727-精简版)》文档中提到的“参考工程文件 C5TB_GHRD.zip”，由于该文档是操作方法的介绍，并没有单独提供相

应参考工程文件，相应的应该使用《车牌识别 bmp 版本 实验指导书》提到的“C5MB_GHRD.....tar.gz”。

8、问题：Quartus QSYS 中的 AccSystem 这个 IP 核是如何添加的
项目 DEMO 中的 AccSystem 这个 IP 核是如何被添加的？是手写的还是通过某种方式自动生成的？

答：AccSystem 是自定义的 IP，是自己写的。

9、问题：参考工程中的 parameter.txt 参数文件如何生成？
parameter.txt 文件有没有具体的生成方法或者参考？

答：parameter.txt 本身没有什么特别的格式约束，是由模型训练所得参数和 C 语言解析所需共同决定的。也就是说 parameter.txt 的数据来自模型训练所得，排列顺序和格式并不重要，只要能在 c 语言里解析提取出来就行。

1、问题：开发板不稳定

在连接开发板并上传 bmp 文件进行识别时，经常到一半出现板子断开情况，请问是因为板子识别图片数量有限还是其他原因呢

答：您好！demo 工程仅供参考，工程设计功能是单张图片识别，不能上传多张图片。需要自行设计程序来适应批量识别。另外如果工程生成的 rbf 文件存在问题也会导致开发板断开情况。如遇到开发板断开情况，可以发邮件到 liuke@awcloud.com，告知开发板 IP 地址。

2、问题：反向训练的平台及过程

可以在本机上使用其他语言（如 MATLAB,python）进行训练吗，视频中要求涵盖训练的过程，那么训练的时间比较长，等待结果的时间需要全部录入视频中吗

答：您好！实现方式这里不做限制，在个人电脑或者云主机都可以，视频可以做剪辑处理加速或者跳过训练中间过程。

3、问题：ARM-EDS 工程

请问这一项具体指的是什么呢

答：比如 demo 工程的 c 工程是用 soceds 开发的（其实就是 eclipse），如果采用其他开发工具则是对应软件的工程文件。

4、问题：平均帧率

要求平均帧率在 5 帧以上具体是什么意思？

答：就是处理一张图片的时间在 200ms 以内（1 秒钟识别 5 张及以上），如果识别速度未达到要求，在初赛阶段不做限制，依旧可以上传作品。

【A21】面向中小型桥梁的健康监测与智能管养系统【法尔胜】

1.问题：关于给定数据问题

请问一下，根据数据所给定的数据：应力、温度、伸缩缝、沉降和索力，这五个数据如何判定危桥？以及这五个数据的具体解释，有何具体作用？这些数据的变化对于各个监测点有什么实际的变化？

答：因为本课题旨在锻炼大家软件开发与数据处理方面的专业技能，因此对于危桥判定可为数据波动设置上下限值，超出限值报警/预警的方式实现，至于限值是否符合桥梁专业中危桥判定标准不在本次课题要求范围内。但从数据归类、判定的角度出发，仍然建议报警/预警限值以可设置/可配置形式输入，开发出来的系统平台通用性会更高。

以上 5 个参数均为桥梁关键技术指标，通俗的解释为：应力主要指桥梁主要构件的内力情况，温度指桥梁构件所处温度场温度情况或环境温度；伸缩缝是桥梁重要构件之一，主要防止热胀冷缩对桥梁产生影响后导致桥梁桥面变形，为桥梁横向伸缩提供空间；沉降主要指桥梁桥墩/桥塔受重力或水土流失、地下地质情况变动等因此的下沉；索力即大型桥梁如斜拉桥、悬索桥钢索受力值。监测数据即通过传感器实际测得的以上部位实际受力/温度/位移变化情况。

2.问题：我们可以使用的开发软件有哪些？

对桥梁各项参数进行模拟仿真，及桥梁模型的建立用什么软件比较好？谢谢指导。

答：开发软件不限，目前三维模型技术可以采用 BIM 技术或者 Unity3D 技术，如果采用 BIM 技术，可以使用 Revit 建立桥梁模型，如果采用 Unity3D 技术，可以使用 3D MAX 进行建模。建议开发环境不要涉及侵权问题，尽量考虑后续平台使用的收费问题。

3.问题：关于人工巡检数据
需要录入的人工巡检数据有哪些

答：人工巡检数据主要是指对桥梁表面如锈蚀、路面开裂、实际产生的结构裂缝宽度等人工巡查获取的图片及测量数据，也包含定期检测中有资质检测单位提供的一些检测报告，检测报告不限内容，因其具体内容根据实际用户需求会有不同。

4.问题：如何判断桥梁健康状态
可以简单的说一说如何去判断桥梁的健康状态吗

答：同问题 1.

5.问题：关于桥梁实体模型
我们需要创建实体桥梁模型吗？或者说是需要安装实体的传感器吗？如果不需要的话，我们只要使用企业提供的那些数据就可以了么？

答：不需要实体模型及实体传感器，仅利用数据即可，可对数据进行适当复制并修改（如修改突变值、异常值等）等以展现更好的平台使用功能及效果。

6.问题：数据融合分析功能中，数据集没有提供人工巡检数据，如何做关联匹配分析和对比验证

答：在实际的桥梁健康监测工程中，实时健康监测与人工巡检数据相结合可以进一步提高桥梁健康状态评价的准确性。这部分没有提供相关数据以供验证，重点也不在此。各参赛队请聚焦在如何实现人工巡检功能，如何设计与实时健康监测互补的人工巡检方案，如果能够模拟出部分示例数据更为理想。

7.问题：桥梁数据

对于企业所提供的数据集，其中长时数据和短时数据是来自同一个桥梁吗？还是来自同一类桥梁？是拉索桥还是拱形桥？还是其他类型的桥？

答：企业提供的数据来源于一座较小型斜拉桥，数据来自于同一桥梁。

8.问题：桥梁长短时预测数据的疑问

长时预测的数据分别是应力监测、温度监测、伸缩缝监测、沉降监测 4 类数据，短时预测的数据分别是应力监测、温度监测、索力监测 3 类数据，为什么不一样啊？

答：企业是从现场的实际监测数据中截取了适合做长时预测和短时预测的数据集，各参赛队可以聚焦在预测算法上，对具体的物理参数值不必过多纠结。

9.问题：题目中要求录入图片信息相关的问题

图片中信息如何提取出来？需要提取出来保存还是直接只保存数据。图片究竟是哪一类图片？文字表格？普通景物图？还是桥梁的细节图？图片有无相关的示例？

答：录入图片一般值桥梁巡检人员现场拍摄的实景照片，示例如下。数据录入主要指录入数据库内的内容包含：图片上传时间、上传人员、对现场情况的描述、裂缝的张开值等，需录入的信息与图片一样，有桥梁巡检人员巡检后手动通过人工巡检功能录入以上信息。

10.问题：需求中的驾驶舱相关问题？

驾驶舱大屏是在哪个平台上进行显示？
有无相关示例文件？

答：主要指大屏演示功能，采用 Web 方式实现即可。为减少对参赛人员的思维限制，并未提供示例文件，大家可网上寻找灵感。演示功能主要目的为在无人为干预、操作的前提下能够但不局限于展示系统重要的信息、主要功能、预报警报及数据变化情况。

11.问题：数据的相关问题

在数据中应力温度和索力在每个检测点上有两列数据，这两列数据是代表同一个检测点在同一时段重复检测两次，还是一个检测点的同一检测特征的不同方面的值，例如温度一个是内表面温度，一个是外表面温度？

答：两列数据是在同一横截面的不同位置监测所得。

12.问题：关于数据

能不能劳请解释一下 表格中各项数据是实测值还是传感器原始数据之类的啊？没有单位也没有说明读的不是很懂...

答：数据是在实测值基础上进行稍许处理获得。数据单位：应力 (MPa)、温度 (°C)、位移/沉降 (mm)、索力 (kN)。

13.问题：预测的相关问题

请问在赛题中提到需要对桥梁的健康检测数据进行预测。请问这个预测是要对每一个检测内容（沉降、伸缩缝等）值再对应桥上的每一个检测点做数据预测吗？

答：需要针对每一项监测内容进行预测，每个监测点的数据也都应该进行预测。

1.问题：人工巡检相关问题

人工检测是和传感器检测同步进行吗？还是人工巡检是在传感器报错后进行的巡检，从而进行数据融合分析

答：人工巡检是周期性的工作，但检测的频率比较低，也不能长期进行。人工巡检和传感器监测有些项目相同，也有很多项目不同，是互为辅佐和补充的关系。

2.问题：监测数据

请问，给出的监测数据中不同表内的标号相同的点检测位置相同吗？例如长时检测温度表中的 A1 点与长时检测索力表中的 A1 点相同吗？长时检测中的点和短时检测中标号相同的点检测位置相同吗？

答：不相同，建议单独考虑长时预测和短时预测的数据，独立进行预测算法的设计。

1、问题：数据预测问题

数据说明中提出“利用提供的数据采集间隔为 1 小时的监测数据集实现 1 个月内的长时预测”，这个意思是不是做间隔一小时的下一个月的长时预测

。然而短时预测中将 1 小时作为一种预测。所以我们应该以一天作为间隔去预测还是以一小时作为预测？

答：应该是按源数据频率进行预测

2、问题：数据预测中某小时时间戳缺失
时间戳缺失是否需要插补。

答：按实际情况来说应该不需要插补

1、问题：桥梁注册信息
桥梁注册需要录入那些信息？

答：桥梁注册信息？似乎并没有涉及到这个概念。

从桥梁的基本信息角度出发，主要还是一个文件导入的功能，另外在后续桥梁运营过程中，会采用人工方式进行测试，有些测试报告文档或者相关数据也可以进行录入，这个录入不只是存档，而是可以录入数据库进行查询等等，如某位置发生裂缝，我可以采用查看自动监测数据曲线一样的操作方式看到裂缝大小随时间的变化情况，当然这个数据来源是录入数据而非传感器数据。

【A22】冰箱智能体验设计【青岛海高】

1.问题：关于聚焦问题的选择
任务清单要求聚焦的问题一定是“问题说明”中的“食材信息录入方式的改进以及针对用户的信息有效输出”，还是可以通过自己的调研聚焦到另外的痛点问题？

答：可以关注其他痛点问题。企业在于获得真实用户需求，所以来源于同学自身的实际体验而锁定的痛点问题以及提供的独特解决方案可能会更有竞争力，不限于问题说明的要求。

2.问题：关于赛题理解方面
贵企如何定义赛题中“80%用户的使用场景”（关系到调研）？

答：这只是二八原理的引用，希望产出更偏向惠及大多数用户的方案，少部分个人喜好容易得到缝隙需求，企业投入产出比不高。

3.问题：关于赛题理解方面

贵企如何定义“智能冰箱”（实现什么程度的智能功能即可算作智能冰箱）？

答：提供物联网相关便利的同时，不要对用户造成反感。具体的，小到对冰箱开关门的检测，通过判定开门关门的状态和时间，帮助用户更好的使用冰箱和管理饮食习惯。

4.问题：设计结果

请问对冰箱的智能设计需要代码实现吗？还是只需要方案设计；赛题中说的迭代过程指的是每一个功能算法实现的过程记录吗？还是可以针对某一个功能进行演示

答：最初级的结果是能提出有竞争力的概念，在此基础上，推荐跨学科组队，合理利用资源，不断完善概念的可实现性，达到更加完整的方案展示。

5.问题：对于选题的方向与方案的问题

（1）该选题是否只要是关于冰箱的创意设计提高用户体验设计都可以？

答：是的

（2）一定要仅限于赛题给的海尔冰箱吗？

答：不局限于此

（3）对于方案的落地，如果是大型冰箱设计，对于外观方面有要求吗？还是只需要 3D 的模拟外观？

答：建议专注于其中某一个方面的概念点做细做精，外观不做太多要求

（4）该赛题是否更侧重于用户体验的原型设计，产品的交互设计与落地实施哪个占的比重更大呢？

答：是的，用户体验会占比稍高些。找到能触动用户的概念点。

6.问题：设计问题

冰箱和手机上的 app 连接，那这个 app 是让我们自己开发吗，交作品的时候是不是还要展示出 app 开发的这个编程过程，还是就只需要我们展示出 app 上的页面是什么样子，具体 app 是可以不存在的

答：这仅关系到作品的完整度。而用户问题挖掘以及优质的创意概念点的本阶段更看重的

1、问题：关于方案规划的问题。

该方案能大屏与模块同时做吗？因为他们有一定必然的联系。

答：可以，但需要考虑在大屏交互的唯一性是否牵强。模块和大屏如果是强关联，那会更具有说服力。

2、问题：关于一个关键问题的选择

如果团队选择针对用户的信息有效输出方面的问题为关键问题，那么其实涉及到食材信息录入的方式我们能否认为已经完备？

答：可以，能把一个方面做深做透也是可以带来用户利益点的，不用分散精力在不必要的方向。

【A23】基于区块链的学位学历认证管理系统【云象网络】

1.问题：请问学位证书是由学校还是教育部上传的？

请问学位证书是由学校还是教育部上传的？

答：只做原型系统，不需要对接学校和教育部系统。

2.问题：请问能否详细讲述证书查询和证书核验的具体功能？
查询需要的数据和返回的结果，以及核验需要的数据和结果

答：只做原型系统，核验证书的哈希值，跟链上的哈希值一致即可。

3.问题：学生是否能查询自己拥有的证书？

学生是否能查询自己拥有的证书？

答：只做原型系统，查询功能可以有。

4.问题: 学位学历的数据

题目中的认证颁发时的数据上链的数据是哪来的? 是公司提供吗, 还是说这个数据可以由我们自己设计

答: 有开发者自行设计。

5.问题: 学位学历系统要涉及到学历的相关功能设计吗

学位一般是指高等学校颁发的证书;

学历是指受教育的经历, 例如小学、初中、高中

那么在设计系统的时候需要设计学历相关功能嘛? 还是说学历学位在本题目中是一个意思, 均指高等学校的证书

答: 不需要涉及学历相关功能涉及。

6.问题: 关于 A23 区块链选择问题

如题, A23 问题中区块链选择能否使用自己的区块链吗

答: 可以。

7.问题: 是否只能采用赛题方给定的技术框架?

能否使用一些别的区块链技术进行实现, 例如 Hypeledger Fabric。

答: 可以

8.问题: make gvnt 报错

```
jiayu@DESKTOP-DPMB6AA:~/go/src/github.com/vntchain/go-vnt$ make gvnt
```

```
build/env.sh go run build/ci.go install ./cmd/gvnt
```

```
>>> /usr/lib/go-1.13/bin/go install -ldflags -X
```

```
main.gitCommit=298e779f11a2cab8da88ec9afaae756cf297463b -v ./cmd/gvnt
```

```
go: finding golang.org/x/sync latest
```

```
go: finding github.com/vntchain/vnt-wasm latest
```



```
go: go-vnt/cmd/gvnt imports
```

```
github.com/vntchain/go-vnt/core imports
```

```
github.com/vntchain/go-vnt/core/wavm imports
```

```
github.com/vntchain/vnt-wasm/disasm: github.com/vntchain/vnt-wasm@v0.0.0-20190111083450-104fc73f682f: parsing go.mod:
```

```
module declares its path as: github.com/go-interpretter/wagon
```

```
but was required as: github.com/vntchain/vnt-wasm
```

```
util.go:45: exit status 1
```

```
exit status 1
```

```
make: *** [Makefile:15: gvnt] Error 1
```

答：是在什么环境下编译的？有可能是包拉取不到，可以根据具体拉取不到的包，先找下对应的仓库，如果找不到，分析下具体怎么替换。（企业老师刚建了个虚拟机，按照<https://github.com/vntchain/go-vnt> 上面的编译，没有遇到什么问题）

1、问题： bottle migrate 出现错误

搭好本地测试网后，使用 bottle init, bottle build 之后，再执行 bottle migrate 出现如下错误

```
fork/exec
```

```
/root/bottle/build/_workspace/src/github.com/vntchain/bottle/build/lib/node/node/bin  
/node: no such file or directory.
```

答：根据错误提示检查一下文件或目录是否存在

2、问题： 认证验证时的验证方法： 核实电子证书文件的哈希值

请问电子证书文件是指代表证书的图片或 pdf 这类文件吗，还是可以由一串数据代表的

答：可以由一串数据代表

3、问题：array 使用问题
array 的 length 在初始化赋值后还可以改吗？
array 支持动态数组形式吗

答：可以，支持动态数组形式

1、问题：赛题侧重点
关于该系统，发包方倾向于算法性能上的优化还是功能完善

答：应用功能开发，不涉及优化

1、问题：vnt-kit 和 bottle 安装失败
vnt-kit 无法连接到服务器，bottle 镜像下载好之后有一个需要下载压缩包无法连接到服务器，下载不了

答：参考一下这个项目 <https://github.com/zjufishboy/PatentX>。建议按上面发的那个 github 上开源的这个项目的步骤做一遍看看，如果安装失败的话，应该是有些模块 github 上下载时国内的网络原因。可以使用 docker 的方式，如果不使用 docker 方式，那要根据下载失败的 url 记下来，在 hosts 文件中给它指定对应的 ip

2、问题：
请问一定要采用 <https://github.com/vntchain> 上的工具开发嘛？

答：参考一下这个项目 <https://github.com/zjufishboy/PatentX>

3、问题：编译器下载好，但是不能初始化合约
问题图片地址

<https://img-blog.csdnimg.cn/20210324145517397.jpg>

https://img-blog.csdnimg.cn/20210324145517682.png?x-oss-process=image/watermark,type_ZmFuZ3poZW5naGVpdGk,shadow_10,text_aHR0cHM6Ly9ibG9nLmNzZG4ubmV0L3N5aXRpbmc=,size_16,color_FFFFFFFF,t_70

https://img-blog.csdnimg.cn/20210324145517684.jpg?x-oss-process=image/watermark,type_ZmFuZ3poZW5naGVpdGk,shadow_10,text_aHR0cHM6Ly9ibG9nLmNzZG4ubmV0L3N5aXRpbmc=,size_16,color_FFFFFFFF,t_70

答：参考一下这个项目 <https://github.com/zjufishboy/PatentX>。 `bottle init` 的那个命令是生成一个脚手架用的，如果这个命令遇到问题，可以不使用 `bottle init`，可以直接用一下上面这个项目的目录结构就好了。