



**中国大学生服务外包  
创新创业大赛**

China Students Service Outsourcing  
Innovation and Entrepreneurship Competition

**第十一届中国大学生服务外包创新创业大赛  
企业命题类赛题手册**

中国大学生服务外包创新创业大赛组委会

二〇一九年十一月

## 目录

一、	概述.....	2
二、	企业赛题.....	4
2.1	【A01】基于绿色出行场景个人碳排放积分系统开发与运营【八维通】 .....	4
2.2	【A02】基于地铁出行场景的小游戏设计【八维通】 .....	7
2.3	【A03】工业物联网——高性能边缘计算轻量级容器管理引擎设计【文思海辉】 ..	9
2.4	【A04】工夹具全寿命智能管理系统（TMS）【捷普】 .....	12
2.5	【A05】基于人工智能的视觉识别技术【文思海辉】 .....	16
2.6	【A06】交通时空大数据分析挖掘系统【东软】 .....	18
2.7	【A07】酒店视觉 AI 解决方案【虹软】 .....	21
2.8	【A08】智·会学习——今目标“智慧校园”场景挖掘与实践应用【今目标】 ..	24
2.9	【A09】僵尸企业画像及分类【科创信息】 .....	26
2.10	【A10】金融科技服务平台企业数据的无监督分类系统【浪潮】 .....	30
2.11	【A11】多方数据安全计算平台的设计与实现【浪潮】 .....	33
2.12	【A12】基于算法的配送路线优化系统【华云+中诺思】 .....	36
2.13	【A13】运用文本相似度实现主观题自动评阅【华云+中诺思】 .....	39
2.14	【A14】金融领域公司实体消歧系统构建【恒生电子】 .....	42
附件一：	A类企业命题初赛统一评分标准（仅供参考） .....	49

## 一、概述

第十一届中国大学生服务外包创新创业大赛（以下简称“服创大赛”或“大赛”）企业命题类竞赛邀请具有代表性的企业参与命题，所有赛题组成赛题池，参赛团队可在赛题池中选择任一赛题参赛。本类竞赛重点考察参赛团队的专业技能及专业竞争力水平。

命题企业根据自己的真实业务需要发布赛题，由参赛团队按要求进行回应。参赛团队需接受参赛承诺书中规定的知识产权条款，赛题涉及特殊知识产权的部分由企业赛题中单独约定。

根据赛题要求不同，部分赛题要求不能使用开源代码的，则选择该赛题的团队禁止使用。如赛题允许使用开源代码，则需要作品中标明哪部分使用了开源代码。对于源代码，大赛不要求参赛团队统一提交，各参赛团队可根据赛题和自身作品情况自愿提交。如若在评审过程中，该参赛项目存在异议或抄袭的嫌疑，评委可要求学生团队提交源代码或要求学生与其进行远程视频答辩。

企业命题类每道赛题不限参赛团队数目。企业命题类竞赛中，除正常比例的一二三等奖，获奖团队可获得与命题企业进行项目对接和成果转化的机会。

本届大赛企业命题类赛题以新一代信息技术为主，主要有以下方向：

1、企业服务——企业服务外包，如法务、财务、人力资源、知识产权、咨询策划、后勤保障、办公协同、软件开发等。

2、智能计算——大数据、云计算、人工智能外包，如数据整理、数据挖掘、数据存储、数据处理、数据应用、数据传输、数据加密等。

3、文创娱乐——内容创意以及消费级内容产品，如动漫、影视、二次元、游戏等。

4、消费互联——面向消费者的应用开发及消费升级产品，如电子商务、品牌设计、消费O2O、大教育、大健康、大消费等。

5、智能制造——分组内容：物联网与工业自动化，如机器人、视觉智能、3D打印、无人机、设备管理、远程控制、新能源等。

6、工业设计——面向工业与硬件产品的设计服务外包，如智能硬件、视觉设计、平面设计、交互界面、集成电路设计等。

## 第十一届服创大赛企业命题类赛题列表

赛题编号	命题企业	题目类别	专业方向	赛题名称
A01	八维通	商业类	消费互联	基于绿色出行场景个人碳排放积分系统开发与运营
A02	八维通	应用类	文创娱乐	基于地铁出行场景的小游戏设计
A03	文思海辉	应用类	智能制造	工业物联网——高性能边缘计算轻量级容器管理引擎设计
A04	捷普	应用类	企业服务	工夹具全寿命智能管理系统（TMS）
A05	文思海辉	应用类	智能计算	基于人工智能的视觉识别技术
A06	东软	应用类	智能计算	交通时空大数据分析挖掘系统
A07	虹软	应用类	智能计算	酒店视觉 AI 解决方案
A08	今目标	应用类	智能计算	智·会学习——今目标“智慧校园”场景挖掘与实践应用
A09	科创信息	应用类	智能计算	僵尸企业画像及分类
A10	浪潮	应用类	智能计算	金融科技服务平台企业数据的无监督分类系统
A11	浪潮	算法（模型）类	智能计算	多方数据安全计算平台的设计与实现
A12	华云+中诺思	算法（模型）类	智能计算	基于算法的配送路线优化系统
A13	华云+中诺思	算法（模型）类	智能计算	运用文本相似度实现主观题自动评阅
A14	恒生电子	算法（模型）类	智能计算	金融领域公司实体消歧系统构建

注：本表中赛题编号为大赛官网报名系统中赛题编号。

## 二、企业赛题

### 2.1【A01】基于绿色出行场景个人碳排放积分系统开发与运营【八维通】

#### 1. 命题方向

消费互联

#### 2. 题目类别

商业类

#### 3. 题目名称

基于绿色出行场景个人碳排放积分系统开发与运营

#### 4. 背景说明

##### 【整体背景】

绿色生活、绿色出行已成为社会共识，地铁、公交等能够有效减少碳排放的公共交通方式已成为百姓出行首选。碳积分(Carbon Credits)指二氧化碳排放配额，企业或个人可通过购买碳积分消除碳足迹。它是基于消费者在出行、住宿、购物等消费过程中，通过采取有效措施，降低碳排放所形成的减排量。

城市私家车是个人碳排放的主要来源，而鼓励人们使用公共交通出行可以极大地促进个人碳排放的降低。因此，如何设计一套完整的运营系统，对个人乘坐公共交通减少的碳排放进行认证、交易，通过物资及经济激励促进城市人群更多地使用公共交通出行，已成为一个具有巨大社会效益和经济效益的课题。

如何结合具体的使用场景与平台特点，利用碳积分有效量化用户绿色出行贡献，转换为用户实际收益，是系统开发的核心内容。通过提高用户环保参与感及实际收益，提高用户粘性，是系统运营的目标。通过环保主题，提升百姓环保意识，展现创业团队的社会责任担当，是系统开发与运营的灵魂。本命题迎合生态文明建设大方向、环保出行大共识、并承载着一定社会责任，商业化运营前景十分广阔。

##### 【业务背景】

八维通科技有限公司是中国中车与中国银联合资成立的轨道交通+互联网行业领先企业。在全国 20 多个城市落地了地铁、公交等交通出行互联网 MaaS(出

行即服务)平台。利用八维通既有的交通出行平台，搭建一套公共出行个人碳排放积分系统，以促进八维通平台用户的活跃，为用户创造新的价值，同时为节能减排贡献力量，是本次命题的初心。

## 5. 项目说明

### 【问题说明】

基于公共出行场景，设计个人碳排放积分运营系统，重点在于如何利用碳排放促进用户活跃、更多地乘坐公共交通工具，如何利用碳排放积分，在运营中设计有趣的玩法，让用户在乘车的同时得到更多乐趣及激励。

同时，个人碳排放积分具有绿色环保属性，以及可交易流通的资产属性，利用这些属性设计一套商业模式，使企业在运营碳排放积分的过程中可以获得额外商业价值，或为其他商业合作伙伴提供商业活动合作平台。

### 【用户期望】

在既有公共出行互联网平台上，设计一套具有可操作性、趣味性、可拓展的个人碳排放积分运营系统，实现 APP 内嵌入应用。

要求：充分利用碳积分可交易流通的资产属性，积分可设计多种活动形式或其他激励措施，如游戏元素、积分道具及交易变现等。

最终目的：增强 APP 对用户的吸引力，提升用户活跃度，提高 APP 用户黏度。

拓展延伸：强化绿色出行理念，鼓励绿色出行，释放碳排放积分的商业价值，宣传公司社会责任意识。

考虑场景特点：公共交通绿色出行，公共环境；高峰时段较为固定，人流量大；站点与周边 POI (Point of interest) 的固定联系等。

考虑用户特征：公共交通客户、游客等，以 20-60 岁年龄段为主。

可结合大数据：用户年龄段、性别、规律乘车站点、路线、时段等用户画像标签。

可结合常见的互联网运营体系进行设计，如用户等级、积分、成就系统等。

## 6. 任务要求

### 【开发说明】

需要考虑场景下并发较高，行车过程中网络状况不佳(高延迟，甚至无信号)、

定位信号不佳的条件。

考虑后台参数配置、数据统计需求。

八维通可提供业务接口，如用户行程状态接口（已进站、已出站、进站站点等）、会员等级积分接口等；可提供用户标签数据查询接口。

需要考虑系统运营的市场推广方案。可以有多样推广方案。

### 【技术要求与指标】

在此项目中，要求详细阐述系统设计，包括整体概念、流程、数值模型；系统架构，与外部接口关系等。

### 【任务清单】

（1）完成项目总体策划，包括商业模式，团队组建，技术方案、流程，市场推广等；

（2）结合企业提供的接口，实现技术对接、程序设计；

（3）给出系统运营方案，包括市场研究与推广。

### 【提交材料】

（1）项目概要介绍；

（2）项目简介 PPT；

（3）项目详细方案；

（4）项目演示视频；

（5）企业要求提交材料：

➤ 商业计划书

➤ 市场推广方案

（6）团队自愿提交的其他补充材料。

## 7. 参考信息

（1）开发工具：开发工具以及开发平台不限，可以借助开源的工具；

（2）数据接口：用户行程状态接口、会员等级积分接口、用户标签数据查询接口。

## 8. 评分要点

本题评分要点参考附件一：A类企业命题统一评分标准。

## 2.2 【A02】基于地铁出行场景的小游戏设计【八维通】

### 1. 命题方向

文创娱乐

### 2. 题目类别

应用类

### 3. 题目名称

基于地铁出行场景的小游戏设计

### 4. 背景说明

#### 【整体背景】

近年来,结合游戏元素提升互联网产品的使用体验与用户粘性是一个被广泛应用的手段。游戏化设计(Gamification)或直接嵌入应用内的小游戏越来越常见,如支付宝蚂蚁森林等,能够给产品运营带来可观的提升。如何结合具体的使用场景与平台特点,设计最合适的游戏元素,达到最佳运营效果,是一个有趣的问题。

#### 【业务背景】

八维通科技有限公司是中国中车与中国银联合资成立的轨道交通+互联网行业领先企业。在全国 20 多个城市落地了地铁、公交出行互联网平台产品。然而,如地铁扫码乘车 APP 等产品具有较强的工具属性,用户停留时间较短,因此需要探索各种运营手段来提升用户活跃与粘性。

### 5. 项目说明

#### 【问题说明】

基于地铁出行场景设计实现 APP 内嵌入小游戏。

考虑场景特点:绿色出行,低碳环保;上下班高峰人流量大、时间固定;站点与周边 POI(Point of interest)的固定联系等。

考虑用户特征:通勤族、游客等,以 20-40 岁年龄段为主。

考虑公共交通的行业特点,游戏主题应健康、积极。

可结合大数据:用户年龄段、性别、规律乘车站点、路线、时段等用户画像标签。

可结合 AR 等新互动技术。



可结合常见的互联网运营体系进行设计，如用户等级、积分、成就系统等。

### 【用户期望】

实现紧密结合场景特点，具有场景特色、趣味性的 APP 内嵌小游戏。

## 6. 任务要求

### 【开发说明】

实现基于 H5 的小游戏设计开发。若需要结合 AR 等技术，可以通过集成 SDK 实现。

需要考虑场景下并发较高，行车过程中网络状况不佳（高延迟，甚至无信号）、定位信号不佳的条件。

考虑后台参数配置、数据统计需求。

八维通可提供业务接口，如用户行程状态接口（已进站、已出站、进站站点等）、会员等级积分接口等；可提供用户标签数据查询接口。

### 【技术要求与指标】

在此项目中，要求详细阐述游戏设计，包括整体概念故事、流程、数值模型；给出系统技术设计，包括系统架构，与外部接口关系等。

考虑高并发、高网络延迟下技术方案。

### 【任务清单】

- (1) 完成游戏设计，包括整体概念故事、流程、数值模型等；
- (2) 结合企业提供的接口能力，实现小游戏 demo；
- (3) 给出参数配置、数据统计方案，如管理后台或通过配置文件实现。

### 【提交材料】

- (1) 项目概要介绍；
- (2) 项目简介 PPT；
- (3) 项目详细方案；
- (4) 项目演示视频；
- (5) 企业要求提交材料：
  - 游戏详细设计
  - 小游戏 demo 程序
- (6) 团队自愿提交的其他补充材料。

## 7. 参考信息

(1) 开发工具：开发工具以及开发平台不限，可以借助开源的工具；

(2) 数据接口：用户行程状态接口、会员等级积分接口、用户标签数据查询接口。

## 8. 评分要点

本赛题评分要点参考附件一：A类企业命题统一评分标准。

## 2.3【A03】工业物联网——高性能边缘计算轻量级容器管理引擎设计【文思海辉】

### 1. 命题方向

智能制造

### 2. 题目类别

应用类

### 3. 题目名称

工业物联网——高性能边缘计算轻量级容器管理引擎设计

### 4. 背景说明

#### 【整体背景】

自从本世纪初云计算模型的出现和广泛应用，云计算已经改变了我们的生活、学习和工作，从贵州到冰岛，全球都可以看到大公司的数据中心。对于亚马逊、微软、阿里和腾讯来说，云计算平台也已经成为非常重要的业务和收入来源之一。但随着物联网时代到来，云计算平台将面临着海量设备接入、海量数据、带宽不够、功耗过高和实时性差等高难度挑战。就目前的带宽水平来说，还无法支持设备到云端之间的数据传输，这使得云计算中心实时返回数据决策也变成了不可能的任务。

#### 【公司背景】

文思海辉物联网事业部在智能制造、智能工厂、智能水务、智能家居、智能办公、智慧医疗、智慧能源、智慧物流、车联网等多领域多行业拥有一系列解决方案与成功案例。

OctoIoT Infrastructure（智能物联网）产品生态通过工业物联网技术，应用无

线方式，实现设备间、设备与应用的互联与集成，对各类工业设备、仪器仪表进行数据的快速采集与传输，并对采集的海量数据进行分析与处理形成专业级的可视化报表，实时展示工业现场生产、反馈运维状态。其还利用数据挖掘、机器学习等 AI 技术自动优化参数及流程，助力制造、电力与能源及仓储与物流等诸多行业进行数字化转型与升级。

### 【业务背景】

在数字经济时代，制造业转型升级速度加快，工业物联网行业迎来发展机遇。其中，工业制造领域的转型升级成为工业物联网发展的重要驱动力，世界各国纷纷发布相关的战略举措，抢占新一轮发展战略机遇。有分析认为，到 2020 年，工业物联网在整体物联网产业中的占比将达到 25%，规模将突破 4500 亿元。边缘计算是微数据物联网中心的网状网络，可在将关键数据传输到中央数据中心或云存储库之前在本地处理和存储关键数据。边缘计算有助于优化云计算系统免受与数据传输相关的中断。

## 5. 项目说明

### 【问题说明】

(1) 工业数据的协议不标准。互联网数据采集一般都是我们常见的 HTTP 等协议，但在业领域，会出现 ModBus、OPC、CAN、ControlNet、DeviceNet、Profibus、Zigbee 等各类型的工业协议，而且各个自动化设备生产及集成商还会自己开发各种私有的工业协议，导致在工业协议的互联互通上，出现了极大地难度。

(2) 大量的工业数据是“脏”数据，直接存储无法用于分析，在存储之前，必须进行处理，而对海量不同类型不同结构的数据进行处理，从技术上又提高了难度。

(3) 原先的工业系统都是运行在局域网中，安全问题不是突出考虑的重点。一旦需要通过云端调度工业之中最为核心的生产能力，又没有对安全的充分考虑，造成损失，是难以弥补的。

### 【用户期望】

- (1) 基于容器技术实现水平扩展运行与边缘计算网关应用程序的框架。
- (2) “微服务”形式实现应用程序（边缘计算网关内）的通用数据发布、

接收、持久化等接口。

(3) 可靠的安全性，提升边缘计算网关的可靠性，实现无法通过边缘计算网关侵入设备网络或传感网络。

(4) 轻量级的容器引擎运行与嵌入式设备。

## 6. 任务要求

### 【开发说明】

需要对项目背景，业务痛点有深入了解和阐述。

需要通过分析业务痛点整理出项目功能列表，并进行 breakdown。

需要针对客户期望进行详细阐述实现效果。

### 【技术要求与指标】

具体开发技术不限，但是不能采用商用模块。所有第三方组件均需要采用开源模块。

### 【任务清单】

完成以下设计：

- (1) 设计轻量级容器引擎；
- (2) 镜像文件管理；
- (3) 容器管理；
- (4) 容器化运行的参数管理（CPU、内存使用限制，端口映射）；
- (5) 容器文件系统管理；
- (6) 容器权限配置；
- (7) 容器的调度配置及管理。

### 【提交材料】

- (1) 项目概要介绍；
- (2) 项目简介 PPT；
- (3) 项目详细方案；
- (4) 项目演示视频；
- (5) 企业要求提交材料：
  - 需求说明书及系统原型图一份
  - 详细系统设计书一份

➤ 实现轻量级容器管理引擎（可参考 Docker 及 Rocket）

(6) 团队自愿提交的其他补充材料。

### 【开发工具与数据接口】

无

## 7. 参考信息

(1) 推荐开发语言：C/C++，Java，Python；

(2) 推荐开发工具：不限；

(3) 推荐开发平台：Linux。

## 8. 评分要点

本赛题评分要点参考附件一：A 类企业命题统一评分标准。

## 2.4 【A04】工夹具全寿命智能管理系统（TMS）【捷普】

### 1. 命题方向

企业服务

### 2. 题目类别

应用类

### 3. 题目名称

工夹具全寿命智能管理系统（TMS）

### 4. 背景说明

#### 【整体背景】

生产设备全寿命周期的数据化管理是智能工厂概念的要求之一，它包含了资产和设备管理的全过程，对设备的采购、使用、维修（轮换）、报废等全过程进行精细化管理，以获得设备寿命周期费用最经济，设备综合性能最高的理想目标。工夹具作为重要生产设备之一，被广泛应用于各类高精度产品的制造过程，直接影响着产品的稳定性和一致性，对提高生产效率和改善工件品质起着关键作用。

#### 【公司背景】

捷普电子（无锡）有限公司成立于 2004 年，是一家提供电子产品解决方案的公司，专门为电子公司和高科技公司提供全面的设计、制造和产品管理服务，向全球各地的客户提供从设计、开发、生产、装配、系统技术支持及到最终用户

分销等优质服务。拥有强大的技术优势、锐意创新的专业态度和尽善尽美的服务精神，许多国际知名企业都是捷普无锡的客户。

### 【业务背景】

由行业特点所决定，捷普生产线上使用的工夹具精度要求极高，随着生产规模的日益扩大，流转于各个 **Workcell**（工作部门）之间的工夹具数量极为庞大，传统的人工记录和追踪方式已逐渐难以满足企业对工夹具的精细化管理要求，企业迫切需要一个全新的系统对每一套工夹具实现全生命周期的电子化、智能化管理。

## 5. 项目说明

### 【问题说明】

公司原来对工夹具的管理是通过纸质记录或 **Excel** 台账的方式进行手工管理的，存在着工作量大，精确度差，数据安全性差，结构化查询不便等诸多缺陷，同时不同部门员工录入的信息难以同步。公司希望开发一个各部门通用的标准化的工夹具管理平台，通过设置不同的 **Workcell**（工作部门）和用户权限来保证数据的安全性。同时，公司希望这个新的管理平台具备一定的智能分析能力，可以通过对大量历史数据的分析和筛选，对工夹具的故障概率进行较为准确的预测。

### 【用户期望】

#### （1）功能综述

- 系统为局域网网站形式。
- 使用登录验证和用户权限控制来确保安全性。
- 系统以 **Workcell**（工作部门）为单位分隔数据，各 **Workcell** 之间数据完全独立且互不可见。
- 对于每一套工夹具，系统从采购入库审批流程→领用/归还→定期检测→报障维修→使用记录→报废手续，整个生命周期都有详细的追踪记录。
- 数据可以多种形式进行查询和追溯，并按用户要求显示报表。
- 数据分析和故障预警功能，能通过构件数据挖掘模型、设计数据分析及故障预警算法等手段，在工夹具即将有较高概率发生故障前，通过 **email** 等方式对相关人员进行预警。

- 当工夹具的理论报废时间即将到达时，提前通过 email 等方式对相关人员进行预警。
- 系统预留工夹具定位显示模块，当公司为重要工夹具安装定位设备后，能实时显示工夹具具体位置。
- 在确保数据可靠性与安全性的前提下，尽量实现流程和表单的自动化。

## (2) 用户权限划分

- **Operator I (初级用户)**：进出库操作、提交报修申请。
- **Operator II (高级用户)**：提交采购入库申请、修改工夹具基础信息、处理报修申请、提交报废申请。
- **Supervisor (监管员)**：创建和修改工夹具类别、处理采购入库申请、处理报废申请。
- **Manager (Workcell 经理)**：对采购入库申请和报废申请进行最终处理。
- **Admin (系统管理员)**：添加或删除用户、更改用户权限。

## (3) 流程概述

- **采购入库流程**：**Operator II (初级用户)** 提交 → **Supervisor (监管员)** 初审 → **Manager (Workcell 经理)** 终审。  
主要内容：申请人、类别代码、物品代码、采购入库单据号、采购入库日期、物品照片。
- **进出库流程**：产线员工申请 → 扫描员工卡确认身份 → 仓管员录入信息 → 工夹具出库/返库。  
主要内容：领用人，操作人，产线，日期，物品代码。
- **报修流程**：**Operator I (初级用户)** 提交 → **Operator II (高级用户)** 处理并关闭。  
主要内容：申请人、物品代码、故障描述、故障照片、处理人。
- **报废流程**：**Operator II (高级用户)** 提交 → **Supervisor (监管员)** 初审 → **Manager (Workcell 经理)** 终审。  
主要内容：申请人、物品代码、物品寿命计数、报废原因。

## 6. 任务要求

### 【开发说明】

使用 Visual Studio 开发一个 MVC 局域网网站，前端通过动态页面向用户进行数据展示并负责与用户互动，后端使用数据库对工夹具数据进行流程化管理。

### 【技术要求与指标】

- (1) 项目使用 B/S 架构；
- (2) 使用关系型数据库设计理念，主键、外键、索引的创建与使用清晰合理；
- (3) 各部门（Workcell）数据共用表格，但在访问上必须完全隔离；
- (4) 前端 UI 需要具备有一定的设计感和美观性；
- (5) 前端 UI 兼容 PC 端与移动端；
- (6) 数据保留 10 年以上。

### 【任务清单】

- (1) 数据库结构设计；
- (2) MVC 项目创建；
- (3) 项目部署与测试；
- (4) 数据库性能优化；
- (5) 数据自动备份的实现。

### 【提交材料】

- (1) 项目概要介绍；
- (2) 项目简介 PPT；
- (3) 项目详细方案；
- (4) 项目演示视频；
- (5) 团队自愿提交的其它补充材料。

### 【开发工具与数据接口】

- (1) IDE 使用 Visual Studio 2015 或更高版本；
- (2) 采用 .NET Framework 平台；
- (3) 系统采用 MVC 架构；
- (4) 后端语言采用 C#；



(5) 前端采用 Bootstrap 和 JQuery 框架;

(6) 数据库采用 MS SQLServer 或 MySQL。

## 7. 参考信息

无

## 8. 评分要点

本赛题评分要点参考附件一：A 类企业命题统一评分标准。

## 2.5 【A05】基于人工智能的视觉识别技术【文思海辉】

### 1. 命题方向

智能计算

### 2. 题目类别

应用类

### 3. 题目名称

基于人工智能的视觉识别技术

### 4. 背景说明

#### 【整体背景】

随着人工智能与各科技领域的深度融合，已经在机器学习、语音识别、计算机视觉等领域产生出了众多创新解决方案。尤其是拥有 80% 人类感知视觉信息的计算机视觉技术与人工智能技术已密不可分。业界人士普遍认为，人工智能发展的下一个阶段，将是计算机视觉技术引领的创新时代。机器视觉技术历经半个多世纪的发展慢慢趋于成熟，正逐步从实验室理论研究走向应用市场，与其相关的应用和任务也逐渐进入到人们的日常生活当中。当前科技进步的速度可以用突飞猛进来形容，可以预见在未来，机器视觉技术会在全世界各项技术全面进步和视觉研究人员的共同努力下达到更高的水平，从而让高可靠性的便捷视觉检测以及视频分析技术直接服务人们的工作和生活。

#### 【公司背景】

文思海辉技术有限公司是中国最大的面向全球客户提供服务的 IT 服务公司，成立于 1995 年，有 20 年丰富的 IT 服务行业经验，全球拥有 26000 余名员工。公司拥有超强的全球运营能力、严格的质量标准和高效的交付流程，致力于成为

全球企业“新时代的合作伙伴”，为您的成功保驾护航。凭借可信赖的交付能力，文思海辉成为 Gartner、The Common Sense Advisory Group、IDC、国际外包专业协议（IAOP）等分析机构公认的行业领导者。文思海辉位列 IAOP “2014 全球外包企业 100 强” 榜单第八名，成为首家跻身该榜单十强的中国企业。2015 年，文思海辉成为中国第二大银行 IT 解决方案供应商。

目前文思海辉物联网事业部关注的领域在物联网方向，细分领域分为工业和民用两大块。同时专注于大数据和 AI 技术的研究和应用，语音，图像作为 AI 中两大应用场景的关键输入，将直接决定未来 AI 在具体应用场景中发挥的价值。该命题取材智能图像识别技术，贴合部门整体战略方向和研究重点。

### **【业务背景】**

工业化环节的人工智能应用，绝大多数都与机器视觉技术有关，客户急切的希望能通过神经网络软件，人的不安全行为是指生产经营单位从业人员在进行生产操作时的违反安全生产客观规律有可能直接导致事故的行为，它们是造成事故的直接原因。对 2017 年全国安全生产数据分析后，发现生产安全事故中 95% 的原因是作业人员的不安全行为导致的，如越权限进入工作场所、违章操作、未正确穿戴劳保用品、操作失误或忽视安全警告等行为。

## **5. 项目说明**

### **【问题说明】**

本题着力于生产环境的安全行为，通过摄像头采集并识别职工在生产车间是否佩戴安全帽的图像数据，开发一个车间安全帽检测系统。企业提供 7581 张含有佩戴安全帽的图片，并对该数据进行了预标注，选手可以自行对数据进行清洗，分成训练集与测试集，以及自行扩充数据集（需提供数据来源）。

### **【用户期望】**

参赛选手结合业务需求，进行算法模型的开发，实现生产环境安全帽佩戴情况的实时检测，并要求开发可以部署到上位机的系统。

## **6. 任务要求**

### **【开发说明】**

需要对设计的应用场景有深入的了解和阐述。

### **【技术要求与指标】**

深度学习框架推荐 Tensorflow 或 PyTorch。开发过程不能采用商用模块。

### 【任务清单】

- (1) 项目立项，确定项目方向和开发方向；
- (2) 确定开发周期，按照计划进行项目开发；
- (3) 按照比赛要求按时提交相应的比赛作品材料。

### 【提交材料】

- (1) 项目概要介绍；
- (2) 项目简介 PPT；
- (3) 项目详细方案；
- (4) 项目演示视频；
- (5) 团队自愿提交的其他补充材料。

### 【开发工具与数据接口】

- (1) 推荐开发语言：Python；
- (2) 推荐开发工具：PyCharm 等。

## 7. 参考信息

企业数据集将通过网盘链接提供。

## 8. 评分要点

本赛题评分要点参考附件一：A 类企业命题统一评分标准。

## 2.6 【A06】交通时空大数据分析挖掘系统【东软】

### 1. 命题方向

智能计算

### 2. 题目类别

应用类

### 3. 题目名称

交通时空大数据分析挖掘系统

### 4. 背景说明

#### 【整体背景】

智慧交通是城市化进程不断向前推进的产物，是城市交通问题日益严峻和土

地资源短缺状况共同作用下的必然结果。随着城市规模的不断发展，相应增加了城市人口数量和汽车保有量，这种形式下，逐渐加剧了不断增长的交通流和有效交通用地之间的矛盾。受多方面因素的影响，城市交通基础设施建设远跟不上交通流增长的步伐，更无法满足人们顺畅出行的需要，而智能交通系统建立的主要目的在于通过增加交通安全、提高交通效益、提高公交效率，减少交通负荷，提高交通决策的合理性来提高城市交通管理水平，充分利用已有的交通资源，满足人们便捷顺畅出行的需求。

### 【公司背景】

东软面向全球市场提供 IT 驱动的创新型解决方案与服务，致力于推动社会的发展与变革，为个人创造新的生活方式，为社会创造价值。公司创立于 1991 年，目前拥有近 20000 名员工，在中国建立了 8 个区域总部，10 个软件研发基地，16 个软件开发与技术支持中心，在 60 多个城市建立营销与服务网络；在美国、日本、欧洲、中东、南美设有子公司。东软是中国第一个上市的软件公司，也是最先通过 CMM5 和 CMMI(V1.2)5 级认证的中国软件公司。

### 【业务背景】

如何方便居民的出行是城市城市规划的重要使命，及时掌握城市居民出行方式及其发展变化，不仅可以及时了解城市当前交通结构状态，同时对于日后涉及城市规划，交通管理等方面也有重要的参考价值。模块的实现，可以提供以人为单位的出行与驻留识别结果，让人的轨迹行为标签化。

城市人口逐渐增多，人流密集的大型活动也日益增多。交通方面，可针对人群密度来分析城市人群的热力分布。

交通出行活动派生于人群的工作或生活目的需求，出行方式的选择其实是交通出行者在特定交通环境下心理和行为的综合反映，影响交通出行者出行方式的选择的因素多且复杂，在其选择出行时刻以及出行路线时，会根据自己的出行经验以及新获取的交通信息做出自己的判断，选择总出行时间最低或者总出行成本最小的路线或者出发时刻。

## 5. 项目说明

### 【问题说明】

以手机信令数据为对象，现有的大数据体系架构和大数据服务组织为基础。

完成交通分析,大数据分析模型和算法的研制,实现人口的出行分析,驻留分析,人群密度分析和出行方式分析等。

#### **【用户期望】**

- (1) 找出信令数据中的潜在价值和规律;
- (2) 利用算法模型,合理实现人口的出行分析和驻留分析;
- (3) 利用分析模型,实现人群出行方式的标签化;
- (4) 了解区域的拥挤程度、高峰时段、是否有异常聚集的现象。

### **6. 任务要求**

#### **【开发说明】**

- (1) 实现数据挖掘、数据调研;
- (2) 实现出行、驻留的分析;
- (3) 实现出行方式分析;
- (4) 实现人群密度(热力)分析;
- (5) 实现前端的展示,实时展示热力图。

#### **【技术要求与指标】**

- (1) 前端技术: JavaScript、Echarts 等前端框架;
- (2) 后端技术: Java、Scala、Kafka、Spark、Zookeeper、Hadoop、Hive、Sqoop、Python 等大数据技术。

#### **【任务清单】**

- (1) 需求调研与分析;
- (2) 系统设计,包含概要设计与详细设计;
- (3) 测试案例编写;
- (4) 编码;
- (5) 部署测试环境测试。

#### **【提交材料】**

- (1) 项目概要介绍;
- (2) 项目简介 PPT;
- (3) 项目详细方案;
- (4) 项目演示视频;

(5) 企业要求提交材料：

- 需求分析文档完整
- 系统设计文档完整
- 测试案例完整
- 算法模型、分析模型完整
- 搭建数据库完整
- 测试报告完整

(6) 团队自愿提交的其他补充材料。

**【开发工具与数据接口】**

大数据平台：自行搭建伪分布式（单机）环境。

**7. 参考信息**

东软提供原始数据（脱敏信令数据）和各模块所需的静态数据。

**8. 评分要点**

本赛题评分要点参考附件一：A类企业命题统一评分标准。

## **2.7 【A07】酒店视觉 AI 解决方案【虹软】**

**1. 命题方向**

智能计算

**2. 题目类别**

应用类

**3. 题目名称**

酒店视觉 AI 解决方案

**4. 背景说明**

**【整体背景】**

随着 AI 技术的不断发展，智能交互已经不断的进入人们的生活。机器变得更加的智能，人们的生活也变得更加的便捷。视觉 AI 技术的应用，不仅仅让机器能够看得见，还能看得懂。如何让 AI 更好的服务于大众，是个非常值得探索的问题。

**【公司背景】**

虹软是计算机视觉行业领先的算法服务提供商及解决方案供应商，服务于世界各地的客户，将领先的计算机视觉技术商业化应用在智能手机、智能汽车、智能家居、智能零售、互联网视频等领域，并且仍在不断探索新的领域与方向。

基于拥有自主知识产权的世界先进计算机视觉技术，虹软已为各领域推出多种计算机视觉解决方案，为全球各类知名的设备制造商提供个性化计算机视觉解决方案。在保持技术领先的同时，虹软率先发布了提供支持离线式图像技术的虹软视觉开放平台，与广大合作伙伴携手推动各类视觉技术应用深入到旅游、教育、政务、出行、社区楼宇、互联网应用等各个领域，引领和推动着视觉技术赋能和落地。

在超过 20 年的发展过程中，虹软成功聚集了众多的视觉领域专家，并吸纳和培养了来自国内外一流高校的优秀人才作为生力军。我们将坚持聚焦在技术，注重技术与行业结合的应用经验，融合先进的学术科研力量，为全球的客户和消费者带来真正价值的视觉享受与体验。

### **【业务背景】**

随着人脸识别技术越来越成熟，人脸识别不断的被用于各行各业中，如银行、居住区、城市、校园等场景。为了更好的让 AI 技术在实际生活中落地，让更多的人能够用上 AI 技术，虹软对外发布了开放平台产品，现已经被各行各业采用。如今在旅游行业，人们生活水平不断的提升，旅游出行或公务出行时，对酒店的服务升级越来越期待。视觉 AI+酒店能否碰撞出火花呢？

## **5. 项目说明**

### **【问题说明】**

利用人脸识别技术，结合酒店场景，打造不一样的酒店服务。可以从酒店入住 Check in、客房进出、酒店用餐、健身房、VIP 等角度思考，结合虹软开放平台人脸技术（人脸识别/人证核验等技术），发挥创新精神，打造不一样的智慧酒店，提升服务水平。

### **【用户期望】**

期望利用人脸识别相关技术，以提升酒店服务体验为核心，形成一个完善的解决方案，并给出原型 Demo。

## **6. 任务要求**

### 【开发说明】

完成酒店视觉 AI 解决方案的设计与原型 Demo 的开发。提供涉及的各端各平台的软硬件演示。

### 【技术要求与指标】

- (1) 开发语言不限。
- (2) 使用一个或多个虹软开放平台免费 SDK 产品。
- (3) 如涉及硬件，则提供 Demo 产品。

### 【任务清单】

- (1) 开展需求分析；
- (2) 提供项目解决方案；
- (3) 完成源码开发；
- (4) 软硬件模拟环境。

### 【提交材料】

- (1) 项目概要介绍；
- (2) 项目简介 PPT；
- (3) 项目详细方案；
- (4) 项目演示视频；
- (5) 企业要求提交材料：
  - 相应的可执行程序
- (6) 团队自愿提交的其他补充材料。

### 【开发工具与数据接口】

(1) 虹软开放平台产品入口：<https://ai.arcsoft.com.cn/index.html>。或 SDK 入口：ArcFace2.2（含人脸检测、比对、跟踪、活体、属性等功能）<https://ai.arcsoft.com.cn/product/arcface.html>。

(2) 人证核验 SDK2.0 <https://ai.arcsoft.com.cn/product/verification.html>。

## 7. 参考信息

无

## 8. 评分要点

本赛题评分要点参考附件一：A 类企业命题统一评分标准。



## 2.8【A08】智·会学习——今目标“智慧校园”场景挖掘与实践应用【今目标】

### 1. 命题方向

智能计算

### 2. 题目类别

应用类

### 3. 题目名称

智·会学习——今目标“智慧校园”场景挖掘与实践应用

### 4. 背景说明

#### 【整体背景】

自 2018 年以来，国家教育部相继出台了《教育信息化 2.0 行动计划》、《中小学数字校园建设规范（试行）通知》等文件，旨在全面加速推进“智慧校园”的建设进程，全面提升校园数字化、信息化、智能化应用水平，力争早日实现“三全两高一大”的发展目标，即教学应用覆盖全体教师、学习应用覆盖全体适龄学生、数字校园建设覆盖全体学校，信息化应用水平和师生信息素养普遍提高，建成“互联网+教育”大平台。努力构建“互联网+”条件下的人才培养新模式、发展基于互联网的教育服务新模式、探索信息时代教育治理新模式，是未来教育发展的必由之路。

#### 【公司背景】

北京今目标信息技术有限公司（简称：今目标）成立于 2005 年，是中国企业互联网行业的领军企业。通过 SaaS 云服务，帮助组织打造一套敏捷、高效、个性化、低成本的信息化管理系统。成立至今，今目标凭借其独特的专利技术，已累计服务了超过 380 万家企业用户。

今目标软件采用模块化设计，20 多款应用可以通过独有的主线功能互通互融，像搭积木一样的搭建符合组织管理与发展需求的业务流程。使用今目标并无定法，不仅可以用于商业领域的企业管理、沟通和协作，凭借系统搭建者聪明的才智与卓绝的设计，同样可以解决个人、家庭、校园生活中的诸多问题。

#### 【业务背景】

纵观市面上的“智慧校园”产品，大多停留在“生活服务”的层面，例如学生考勤、选课、成绩查询、校园一卡通等。但却鲜有以提高协同管理效率、提升学生自我管理水平和改善教学科研水平层面出发的“智慧校园”解决方案。

本赛题的创新点就在于开始思考如何利用信息化工具，切实地帮助学生提升自我管理水平和帮助老师提高传道授业效率和效果，乃至整体提升高校教学科研水平，这些都是建设“智慧校园”的终极思考。

## 5. 项目说明

### 【问题说明】

场景包括但不限于：高校学生个人管理、时间管理、社团活动、老师教学、办公、寝室管理等。请结合校园生活中的管理协同的痛点难点，基于今目标产品，提供解决和实施方案。

### 【用户期望】

基于今目标云平台发掘实践应用场景，提供具体应用案例，围绕应用场景或某一个特定痛点，制作 30 秒-5 分钟的 vlog 小视频进行场景演示。有能力的团队可考虑提供实践应用。

## 6. 任务要求

### 【开发说明】

今目标采用模块化设计，在软件内自主搭建即可，不需要二次开发。

### 【技术要求与指标】

不需要编程能力，简单的学习今目标使用技巧即可。

### 【任务清单】

- (1) 进行前期走访与调研，寻找校园生活、学习、管理中的难点痛点；
- (2) 使用今目标搭建相应的业务场景。

### 【提交材料】

- (1) 项目概要介绍；
- (2) 项目简介 PPT；
- (3) 项目详细方案；
- (4) 项目演示视频；
- (5) 企业要求提交材料；

➤ 演示账户，即学生通过今目标搭建的系统

(6) 团队自愿提交的其他补充材料。

### 【开发工具与数据接口】

“今目标”电脑端软件，下载、安装、注册后即可使用。

## 7. 参考信息

(1) 今目标官网：[www.jingoal.com](http://www.jingoal.com);

(2) 今目标公众号：微信搜索“今目标”订阅号，回复“外包大赛案例”获取往届优秀场景案例。

## 8. 评分要点

本赛题评分要点参考附件一：A类企业命题统一评分标准。

## 2.9 【A09】僵尸企业画像及分类【科创信息】

### 1. 命题方向

智能计算

### 2. 题目类别

应用类

### 3. 题目名称

僵尸企业画像及分类

### 4. 背景说明

#### 【整体背景】

在互联网逐渐步入大数据时代后，可以更便捷的获得企业的工商、司法、经营、上市、知识产权、舆情等多维度数据。2016年9月，国家工商总局发布《工商总局关于新形势下推进监管方式改革创新的意见》（工商企监字〔2016〕185号），指出要“依托大数据加强监管”，“充分发挥大数据在制定完善新型市场监管制度和政策中的作用，搜集掌握经营者、消费者和社会公众的反应，跟踪监测有关制度和政策的实施效果”，“在工商登记、企业监管、网络交易、竞争执法、消费维权等领域率先开展大数据示范应用”。通过对企业从业人数、成立年限、注册资本、营业收入、风险信息，行政处罚、纳税信用等级、黑名单、上市信息、电商信息等数据关联处理，按需进行权重分割，并对这些数据进行综合分

析，构建出企业全息画像，整体评估一个企业综合价值，更好的勾勒出企业的经营变化情况。

如何根据企业全息画像，面向不同的应用场景对企业分类识别，分析挖掘法人、自然人、物品（包括特种设备、产品、食品等）、案事件之间的显性和隐性关系，建立市场主体信用管理与风险评估模型，对企业的信用风险、经营风险、质量风险、安全风险、关联风险等进行全面评估及分级预警，为系统监管、精准监管、高效监管提供数据支撑，越来越受到市场监管部门的重视。

### **【公司背景】**

湖南科创信息技术股份有限公司成立于 1998 年，是国内智慧政务及智慧企业领域的信息化综合服务提供商，主要致力于为政企客户提供集软件开发、系统集成、IT 运维等于一体的信息化综合解决方案。

科创信息注重跟踪最新技术发展趋势，在大数据、云计算、移动互联网、人工智能等领域都已创建了多个具有自主知识产权且处于国内领先地位的核心技术平台。在此基础上，按照“平台+应用”的研发模式，针对政府、企业客户的典型应用需求，逐步形成了“智慧政务”与“智慧企业”系列解决方案，客户覆盖湖南、湖北、云南、河南、北京等多省（市）党政机关及其下属公安、财政、税务、环保、医疗、教育等各级政府部门，并为中国铁路总公司、中国移动、中国联通、中国电信、中联重科、云南冶金、中建材集团等不同行业领域龙头企业提供了多项信息化综合解决方案。

依托强大的研发实力与突出的项目实施效果，公司解决方案及产品曾获得湖南省科技进步奖、国家重点新产品、中国优秀测绘工程奖、中国工业软件杰出贡献奖、中国十佳电子政务解决方案奖、中国智慧政务领域最佳解决方案奖、最佳智慧城市解决方案提供商奖等奖项，公司先后荣获国家高新技术企业、中国服务外包成长型 100 强企业、2017 中国最具影响力软件和信息服务企业等诸多荣誉，是湖南省最具成长性的非公有制企业。

公司秉承“用户至上，品质优先”的服务理念，凭借多年的项目实施经验及良好的服务信誉，已在智慧政务及智慧企业领域形成了自己独特的竞争优势，并取得了良好的社会效益及经济效益。

### **【业务背景】**

僵尸企业是指缺乏盈利能力却能够以低于市场最优利率成本获得信贷资源，依靠外界输血而缺乏自生能力的企业。僵尸企业的存在破坏了市场机制，加剧了信贷资源的错配，带来了严重的产能过剩问题，还对其他非僵尸企业产生了投资挤出效应。此外，由于我国金融制度不健全，僵尸企业往往会倾向于采用企业间商业信用的非正式金融方式融资，带来对僵尸企业合作关联企业的“传染”，甚至会拖累银行成为“僵尸银行”。僵尸企业如果不能及时处置，经济动能转换以及高质量发展就无法保障，去杠杆工作的受阻将加大系统性金融风险。鉴于僵尸企业的危害性，自 2015 年以来，中央出台了僵尸企业处置的一系列政策指导文件，试图以处置僵尸企业作为深化供给侧结构性改革、调整经济结构以及促进经济行稳致远的重要抓手。

僵尸企业的识别标准主要有官方标准、CHK 标准，FN-CHK 标准以及各类 FN-CHK 修正标准等，官方标准定义为不符合国家能耗、环保、质量、安全等标准，持续亏损三年以上且不符合结构调整方向；已停产、半停产、连年亏损、资不抵债要靠政府补贴和银行续贷维持经营的企业。CHK 标准的核心是企业是否接受信贷补贴，FN-CHK 标准则包含“真实利润原则”以及“常青贷款原则”，相关的 FN-CHK 修正方法大部分都是对以上两类标准的修正，对企业利润与资产负债率等指标进行调整，将企业的经营管理费用、净资产水平、企业效率和创新等指标引入僵尸企业的识别标准体系中，力求从更加多维的层次反映僵尸企业的经营特征。

通过各类标准分类识别僵尸企业均有其局限性，尤其是标准的制定有其主观性，因此基于企业画像分类识别方法成为首选。

## 5. 项目说明

### 【问题说明】

命题单位已经多角度、多层次、多领域汇聚了约不同行业的 5 万家企业相关信息，其中定义为僵尸企业的有 2 千家。本任务的数据包括训练集，开发集和测试集，前期将发布训练集和开发集，测试集将在此后公布，直接用于模型评测。

企业数据一部分为通用的标签信息，其他的新增标签由参赛人员根据经验和数据本身特点，针对具体问题制订；然后通过数据治理，并采用统计、分析、挖掘等技术，完善企业画像的各维度标签；最后采用人工智能算法建立僵尸企业的

分类模型，达到对企业进行合理的分类识别的目的。

### 【用户期望】

要求：

- (1) 多表数据融合的能力；
- (2) 数据特征提取的能力；
- (3) 数据预处理的能力；
- (4) 数据建模的能力。

目标：追求模型分类识别的精确性与高效性。

## 6. 任务要求

### 【开发说明】

需要对僵尸企业的表现形式有深入的了解和阐述。

需要依据用户期望的内容对数据的处理过程进行详细的描述。

### 【技术要求与指标】

具体开发技术不限，建议使用 Python 和 R 语言等。

### 【任务清单】

- (1) 模型的概述与简介，对算法模型有充分的认识理解，切不可只会套用；
- (2) 模型的参数调优过程（如果包含调优过程）；
- (3) 模型的效率（包括数据预处理的效率与打标签的效率）。

### 【提交材料】

- (1) 项目概要介绍；
- (2) 项目简介 PPT；
- (3) 项目详细方案；
- (4) 项目演示视频；
- (5) 企业要求提交材料：
  - 所使用的分类方法原理与参数调优过程概述
  - 训练后的分类方法，以及对测试集打标签结果
  - 可视化 web 端开发，支持单个与批量输入查询分类结果
- (5) 团队自愿提交的其他补充材料。

## 7. 参考信息

(1) 开发工具：可以借助开源的工具；

(2) 数据接口：企业会通过网盘提供。

## 8. 评分要点

本赛题评分要点参考附件一：A类企业命题统一评分标准。

## 2.10 【A10】金融科技服务平台企业数据的无监督分类系统

### 【浪潮】

#### 1. 命题方向

智能计算

#### 2. 题目类别

应用类

#### 3. 题目名称

金融科技服务平台企业数据的无监督分类系统

#### 4. 背景说明

##### 【整体背景】

机器学习算法根据训练的学习方式可以划分为监督学习、无监督学习、半监督学习以及强化学习，监督学习中当预测的输出变量属于一系列类别时则为分类，预测的输出变量是实数时则属于回归；无监督学习则使用无标记的训练数据，聚类方法为无监督学习的典型方法。金融科技为机器学习、大数据、人工智能等技术在金融场景领域中的应用，金融科技的应用克服了传统金融信贷场景审核流程长、放贷慢的弊端，应用机器学习方法自动评估小微企业信用水平、企业还款能力等是金融科技在金融场景中的重要应用之一。金融场景中企业这一信贷主体的数据覆盖互联网、政府、线上应用等来源的方方面面，数据量大，来源广泛、涉及企业的维度丰富，在分析企业还款能力、信用水平过程中面临巨大的挑战。本赛题金融科技服务平台企业数据无监督分类分析的目的是为了构建企业画像形成每个企业的明显标签，本赛题题旨在寻找一种有效的无监督分类方法，能够对企业主体进行快速准确的分类，降低人工成本及出错率，并能够形成企业群体的有效簇划分，为后期针对企业进行客群分析、精准获客等应用提供基础。

##### 【公司背景】

浪潮卓数大数据产业发展有限公司作为浪潮集团旗下的大数据板块，致力于成为数据资源提供商、数据资产运营商和数据交易服务商，以大数据时代的“数商”为发展目标，促进数据社会化。

### **【业务背景】**

“一贷通”是公司“一平七通”发展战略中的“一通”之一，“一贷通”的业务目标旨在积极采用大数据、机器学习、人工智能等先进的金融科技技术，汇聚各个政府委办局、区域内金融机构、互联网等多渠道的基础数据，搭建一涵盖各金融业务数据的多功能的数字金融科技服务平台，形成科学、客观、可靠的小微企业信用评分体系，着力解决中小企业“融资难、融资贵”的问题。企业多源数据、多维度的深入挖掘是为企业构建企业画像、建立企业信用评分体系的前提基础，从企业的企业背景、经营能力、经营风险、发展状况等层面对企业进行群体划分，企业划分结果中的每一个企业簇群体都要形成较明显的标签标示，为后续企业画像构建、企业信用评分体系构建提供辅助。

## **5. 项目说明**

### **【问题说明】**

以某一地市的小微企业为研究对象，以该地市小微企业覆盖企业背景、企业稳定性、企业经营能力、企业经营风险、司法风险、信用风险等多个方面的数据作为数据来源。建立一种无监督的分类模型，利用小微企业包含的特征维度信息，对小微企业进行簇划分，划分的每一个簇都有有效的特征或者标签去描述该簇的特征，每个簇之间形成较为明显的划分界限，即最终形成企业合理的划分。

### **【用户期望】**

追求企业无标识脱敏数据的有效划分及每个簇划分标签的合理有效且可区分：

(1) 针对无标识的企业数据进行数据预处理，特征筛选，特征提取等形成有效的训练样例及特征；

(2) 针对提取的有效特征选择合适的无监督分类方法对小微企业数据进行分类，进行模型训练，模型要求实现小微企业群体的有效划分；

(3) 针对小微企业划分后各簇提取显著标签进行该簇的描述，要求标签合理且有效；



(4) 企业无监督分类要求最终以完整系统的形式接收企业信息输入，展示企业划分簇类别、该企业所在簇的有效标签。

## 6. 任务要求

### 【开发说明】

无监督分类是数据科学研究的一个重点课题。金融行业积累了大量的企业脱敏数据信息，企业的有效划分及标识在企业信用评估、企业风险监测中具有重要作用并受到各大平台的重点关注。本次拟定通过对不带标识企业脱敏数据进行无监督训练，建立合适的无监督分类模型方法，对不带标识的企业进行有效、合理的划分。

### 【技术要求与指标】

在此项目中，要求详细阐述使用的数据预处理方法，训练过程中如何对模型进行适配调优，如何选取合适的无监督分类方法，训练所得无监督分类模型的评估指标，最终模型的效率（如为企业划分簇所用时间，最终形成几类有效合理的簇划分，每类簇明显的特征标识）等。

如下列出项目中所需要的部分评估指标：

- (1) 模型训练所需时间；
- (2) 模型预测新企业样本所需时间；
- (3) 模型评估指标，例如准确率，召回率，F1-score 指标，AUC 指标，ROC 指标；DB 指数、CHI 指数、轮廓系数等；
- (4) 系统接收输入企业信息，输出企业划分类别所需时间。

### 【任务清单】

- (1) 模型的概述与简介，对算法模型有充分的认识理解，切不可只会套用；
- (2) 模型的参数调优过程（如果包含调优过程）；
- (3) 模型的效率（包括预处理数据的效率、聚类以及打标签的效率）。

### 【提交材料】

- (1) 项目概要介绍；
- (2) 项目简介 PPT；
- (3) 项目详细方案；
- (4) 项目演示视频；

(5) 企业要求提交材料:

- 所使用的无监督分类方法原理与参数调优过程概述
- 训练后的无监督分类方法, 以及对训练样例进行簇划分后各簇的打标签结果
- 可视化 web 端开发, 支持单个与批量输入查询分类结果

(6) 团队自愿提交的其他补充材料。

## 7. 参考信息

- (1) 开发工具: 开发工具以及开发平台不限, 可以借助开源的工具;
- (2) 数据接口: 企业数据会通过网盘提供。

## 8. 评分要点

本赛题评分要点参考附件一: A 类企业命题统一评分标准。

## 2.11 【A11】多方数据安全计算平台的设计与实现【浪潮】

### 1. 命题方向

智能计算

### 2. 题目类别

算法(模型)类

### 3. 题目名称

多方数据安全计算平台的设计与实现

### 4. 背景说明

#### 【整体背景】

随着大数据、人工智能等数字经济时代新兴技术日益成熟, 在日常生活中已经得到广泛应用, 各行各业沉淀下来的数据背后所蕴含的潜在价值越来越受到大家的高度重视, 数据已成为企业和国家具有战略价值的核心资产。

数字经济时代, 越来越多的企业或组织需要与产业链上下游业务伙伴在数据流通和交易领域进行深度合作。因为只有通过各方数据协同计算, 才能更好地释放数据更大的价值, 提升生产效率, 推进产业创新。数据共享和流通将成为刚性业务需求。

保证个人信息、商业机密或独有数据资源等隐私信息在数据处理、流转过程

中不会泄露，是企业或组织参与数据共享和流通合作的前提条件。

### 【公司背景】

浪潮卓数大数据产业发展有限公司作为浪潮集团旗下的大数据板块，致力于成为数据资源提供商、数据资产运营商和数据交易服务商，以大数据时代的“数商”为发展目标，促进数据社会化。

### 【业务背景】

“一贷通”是公司“一平七通”发展战略中的“一通”之一，“一贷通”的业务目标旨在积极采用大数据、机器学习、人工智能等先进的金融科技技术，汇聚各个政府委办局、区域内金融机构、互联网等多渠道的基础数据，搭建一涵盖各金融业务数据的多功能的数字金融科技服务平台，形成科学、客观、可靠地中小微企业信用评分体系，着力解决中小企业“融资难、融资贵”的问题。企业多源数据、多维度的深入挖掘是企业构建企业画像、建立企业信用评分体系的前提基础，从企业的企业背景、经营能力、经营风险、发展状况等层面对企业进行群体划分，企业划分结果中的每一个企业簇群体都要形成较明显的标签标示，为后续企业画像构建、企业信用评分体系构建提供辅助。

## 5. 项目说明

### 【问题说明】

假如当前有 A、B、C 三个政府数据源，一贷通接入政府数据时需要保证政府数据的安全使用，要求明文的政府数据始终保留在政府内部，这就需要将政府数据加密后输出到一贷通的风控模型中，风控模型对加密数据进行计算，对最终的结果加密数据进行解密拿到明文结果，并且结果和明文计算的结果要保持一致。

### 【用户期望】

对明文数据加密后进行计算，将得到的密文结果解密拿到明文结果，结果和明文数据直接进行计算保持一致。

## 6. 任务要求

### 【开发说明】

在政府数据源端需要部署数据加密端，将数据按特定算法加密后输出。核心计算端负责部署风控模型，风控模型对加密后的数据进行计算，得到最终的评估结果。管理端提供平台的管理功能。

### 【技术要求与指标】

项目中一共包括数据加密端、核心计算端、管理端 3 个模块，开发语言不限，主要用到同太加密相关算法技术，实现数据先加密后计算。

测试说明：

以个人授额评估运算模型为例，个人授额评估模型的明文运算公式  $\text{Math.max}(\text{Math.min}(\text{Math.max}(A * 0.8, B * 15) * 10, 150000), 50000)$ ,

A:个人公积金月缴额

B:个人社保缴费基数

个人授额评估模型的输入项 A、B 分别为加密后的密文数据，系统基于密文数据做上述公式的计算，得到加密后的运算结果，将密文运算结果解密后得到明文计算结果。

### 【任务清单】

- (1) 数据加密端开发，实现数据加密输出；
- (2) 核心计算端开发，实现风控模型的部署，风控模型程序可以对加密数据进行计算，输出结果数据；
- (3) 管理端对平台提供可视化的管理功能，主要功能菜单：
  - 平台的登录注册
  - 风控模型管理，可以管理已经安装的风控模型
  - 数据使用记录，可以查看模型已经使用的数据明细

### 【提交材料】

- (1) 项目概要介绍；
- (2) 项目简介 PPT；
- (3) 项目详细方案；
- (4) 项目演示视频；
- (5) 团队自愿提交的其他补充材料。

### 【开发工具与数据接口】

开发语言及开发工具不限

## 7. 参考信息

可搜索多方计算、同太加密相关资料。

## 8. 评分要点

本赛题评分要点参考附件一：A类企业命题统一评分标准。

## 2.12 【A12】基于算法的配送路线优化系统【华云+中诺思】

### 1. 命题方向

智能计算

### 2. 题目类别

算法（模型）类

### 3. 题目名称

基于算法的配送路线优化系统

### 4. 背景说明

#### 【整体背景】

随着信息技术的发展，现代物流作为“第三个利润源泉”是一种先进的组织方式和管理技术，已被世界各国广泛采用，并形成产业化，在国民经济中发挥越来越重要的作用，这一管理技术，正受到日益广泛的重视，并面临巨大的发展机遇，在现代物流中，配送是一个重要的与消费者直接相连的环节，一方面可以体现企业的核心竞争力，另一方面通过线路优化，可以提高企业的运作效率，降低配送成本，实现物流科学化。目前市场上仓储类管理系统已经比较完善，但对于不同需求的路线规划还存在一定的短板，基于当前站好“最后一岗”的要求，引用路线优化算法，开发出具有实用性，先进性，高可靠性的路线规划系统。

#### 【公司背景】

华云数据集团连续四年被评为中国云计算、大数据独角兽。深圳市中诺思科技股份有限公司是以物联网、大数据为技术核心，提供信息化建设为主的高新技术企业，多年来双方在行业产品开发和研究有众多合作，尤其在物流、大数据相关专业的教育类产品已经比较成功，但目前公司的主要软件类产品主要集中在仓储管理运用方面，运输类的产品相对较少，也相对比较简单，在长期的与客户合作过程中，针对运输配送类产品的需求量也日益增加，基于公司未来的业务发展方向及产品类型智能化的扩充，特选编写此命题进行共同探讨。

#### 【业务背景】

华云数据集团与深圳中诺思科技股份有限公司多年来在教育产品开发和研究有众多合作，尤其在物流相关专业的教育类产品已经比较成功，但目前公司的主要软件类产品主要集中在仓储管理运用方面，运输类的产品相对较少，也相对比较简单，在长期的与客户合作过程中，针对运输配送类产品的需求量也日益增加，基于公司未来的业务发展方向及产品类型智能化的扩充，特选择此命题进行共同探讨。

## 5. 项目说明

### 【问题说明】

通过实际案例描述，根据配送点和业务需求量，进行最优路线的计算。

由单一起点出发，配送多个终点后再回到起点，根据车辆数量，承载限制，不同车辆服务成本、运行里程限制等条件选择最优运输路径，使成本最小化，配送订单最大化，满载率最大化（如由一个配送中心向各个销售点配送货物，通过算法确定配送中心每辆车的配送方案，包括配送至哪个客户，配送量，下一个配送目的地）。

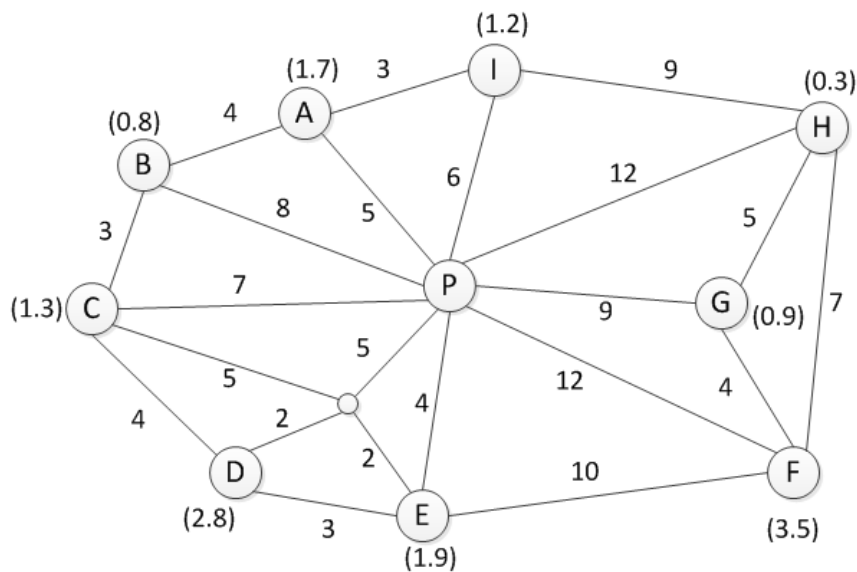
### 【用户期望】

通过算法（蚁群算法、遗传算法等，可任意选择）模型构建，利用计算机语言进行开发（不限定开发语言），最终可通过设置配送点坐标或距离，各配送点业务量，车辆装载等参数，程序能自动计算出最优路线顺序。

## 6. 任务要求

### 【内容说明】

现有一配送网络，图中 P 为配送中心，其余 A-I 为客户的接货点，所有客户的位置距离固定，各边上的数字为公里数，括号内的数字为需输送到各接货点的货物量，单位为吨。假设该配送中心有最大装载重量为 2 吨和 5 吨的两种货车，并限制车辆一次运行路线距离不超过 35 公里，每个派送点只由一辆车服务一次，每辆车只能服务一条路线，车辆一律由配送中心出发，完成任务后返回配送中心，快递车辆配送过程中无装货，只考虑卸货。每个点卸货时间固定为 5 分钟，车辆每小时行驶距离为 10 千米，每个派送人员工作时间为 8 小时，请参考内容信息通过算法和程序计算需要的车辆数，最优路径及配送时间。



根据以上案例，依次轮推，当配送点增加或减少，各配送点之间距离，货物需求量，配载车辆装载量变化时，该算法及程序依然适用。

### 【技术要求与指标】

开发要求：

- (1) 可设置配送点个数，设置网点和各配送点及配送点与配送点之间的距离；
- (2) 设置各网点的货物量，车辆数量、车辆巡回里程限制，车辆的承载量，装卸时间，配送时长等参数；
- (3) 根据变量值系统可以给出最优解决方案；
- (4) 必须采用一种算法应用于程序中。

### 【任务清单】

- (1) 系统架构设计，形成详细设计文档；
- (2) 程序开发；
- (3) 系统测试及部署；
- (4) 用户操作手册。

### 【提交材料】

- (1) 项目概要介绍；
- (2) 项目简介 PPT；
- (3) 项目详细方案；

(4) 项目演示视频；

(5) 企业要求提交的材料：

- 需求分析文档
- 系统设计文档
- 测试案例
- 测试报告
- 系统部署手册
- 用户操作手册

(6) 团队自愿提交的其他补充材料。

#### 【开发工具与数据接口】

不限

#### 7. 参考信息

无

#### 8. 评分要点

本赛题评分要点参考附件一：A类企业命题统一评分标准。

除此之外，企业还会重点关注算法创新（算法创新：算法能够解决同类问题，同时融入一些创新算法设计，具有技术先进，操作说明详细）。

## 2.13 【A13】运用文本相似度实现主观题自动评阅【华云+中诺思】

#### 1. 命题方向

智能计算

#### 2. 题目类别

算法（模型）类

#### 3. 题目名称

运用文本相似度实现主观题自动评阅

#### 4. 背景说明

##### 【整体背景】

AI 时代下，取景翻译，智能分词已不是新鲜的技术，初高中英语作文的自



动化批改，也在一些线上学习系统中有所应用，但对于长篇大段复杂中文语义分析，仍未有较好的解决方案推出，对于大量应用主观题的公考、职业教育、资质类考试，主观题部分目前依旧依靠老师人工阅卷评分。

### 【公司背景】

华云数据集团连续四年被评为中国云计算、大数据独角兽。深圳市中诺思科技股份有限公司是以物联网、大数据为技术核心，提供信息化建设为主的高新技术企业，多年来双方在行业产品开发和研究有众多合作，尤其在物流、大数据相关专业的教育类产品已经比较成功，但随着教育越来越鼓励学生对知识的综合运用，而非单纯的知识水平考试，客观题的考察占比越来越低，方案设计等主观题的考察占比越来越高。传统的考评软件，仅支持客观题自动评分，主观题仍需要老师人工打分。如何将老师从批改试卷、作业这类重复性工作中完全解放出来，特编写此命题进行共同探讨。

### 【业务背景】

随着职业教育越来越鼓励学生对知识的综合运用，而非单纯的知识水平考试，客观题的考察占比越来越低，方案设计等主观题的考察占比越来越高。传统的考评软件，仅支持客观题自动评分，主观题仍需要老师人工打分。在教学智能化日益提升的今天，函需一款软件将老师从批改试卷、作业这类重复性工作中完全解放出来。

## 5. 项目说明

### 【问题说明】

与开放性题目不同，主观题虽然不限用户的答题内容，但仍要求用户能回答到几个关键得分点。而由于中文的语义多样性，对于关键得分点的表述，仍可能存在多种文字描述，而即便出现关键得分点，如果语义完全错误，仍不能认为表述得分，所以需要用户的回答语义进行正确识别，并对关键词进行匹配。学生可以针对提供的部分测试样例做训练，本实验提供的测试样例均来自教师对于学生的真实评分。

如以下示例：

问题描述：物流系统由什么组成？

答案描述：物流系统是由物流作业系统和支持物流信息流动的物流信息系统

两部分组成。

答案关键字：物流作业系统、物流信息系统

回答方式：

物流作业系统和物流信息系统 满分

物流作业和物流信息 满分

不是物流作业和物流信息 0分

作业和信息 50%

物流作业 50%

物流管理、物流采集、物流统计 0分

### 【用户期望】

(1) 提供一个可推广的长文本语义相似度计算方案，提出的方案可广泛应用于不同题目；

(2) 分析语句的主谓宾结构，识别语句中的否定词；

(3) 能够成功完成动词、形容词、副词等的双重否定或反义词的否定与原词的相似性匹配；

(4) 能够针对常见的解释性语句做词语及其解释之间的匹配，如：地球（一个天体）；地球是一个天体；地球：一个天体；地球-一个天体；地球，一个天体；

(5) 在题目判分中，需实现可按回答关键词命中且符合答案中的释义按点给分；

(6) 以试题为例，实现以上描述的主观题自动判分（文本分析试题由企业提供）。

## 6. 任务要求

### 【开发说明】

根据开源的汉语语料库，如 hanLP, jieba 等，对标准答案与学生答案做分析，设计一套比较标准答案与学生答案之间相似度的通用算法，使输入一系列标准答案、标准答案关键字、学生时，给出学生答案与标准答案之间的相似性，考虑每道简答题为 10 分，设计一套合理的评价机制，为学生的答案做评分，注意评分只能为整数且在 0 到 10 之间。

### 【技术要求与指标】

- (1) 能够阐明自己设计的系统的评分机制；
- (2) 考虑算法的效率，包括时间和空间性能；
- (3) 针对测试样例（部分提供给学生），统计参赛人员设计的算法得到的学生各题分数与教师提供分数之间的绝对差值，所有学生的各题分数绝对差值最小者为优。测试为开放性测试，不仅限于提供的样例数据。

### 【任务清单】

- (1) 请参赛者从物流专业职业教育角度，设计主观题自动评阅设计方案；
- (2) 此方案及具体功能均归参赛队伍所有，参赛者最终需要实现并演示该解决方案；
- (3) 方案必须在算法上有贡献，规则占比过大的视为无效。

### 【提交材料】

- (1) 项目概要介绍；
- (2) 项目简介 PPT；
- (3) 项目详细方案；
- (4) 项目演示视频；
- (5) 团队自愿提交的其他补充材料。

### 【开发工具与数据接口】

无要求

## 7. 参考信息

编程语言不限。

## 8. 评分要点

本赛题评分要点参考附件一：A类企业命题统一评分标准。

除此之外，企业还会重点关注算法创新（算法创新：算法独特、新颖，创新元素多，具有技术含量，可以有效地解决此类问题）。

## 2.14 【A14】金融领域公司实体消歧系统构建【恒生电子】

### 1. 命题方向

智能计算

## 2. 题目类别

算法（模型）类

## 3. 题目名称

金融领域公司实体消歧系统构建

## 4. 背景说明

### 【整体背景】

互联网时代，信息爆炸。面对海量资讯，我们希望前沿 AI 技术能将文本和海量实体（公司、人名等）信息关联起来，提升用户阅读流畅度、实现精准化内容推荐等。智能资讯处理，不仅给金融提供智能化服务，还能为金融业务提供更多创新空间。

### 【公司背景】

恒生电子股份有限公司（简称恒生公司）于 1995 年成立于杭州，2003 年在上海证券交易所主板上市（代码 600570），是中国领先的金融软件和网络服务供应商。恒生聚焦于财富资产管理，致力于为证券、期货、基金、信托、保险、银行、交易所、私募等机构提供整体的解决方案和服务，为个人投资者提供财富管理工具。目前，恒生公司拥有约 8000 人的高素质专业队伍，其中研发工程人员约占 80%，在规划咨询、软件开发实施、技术服务、系统集成及系统维护服务等方面，具备强劲实力和竞争优势。多年来，恒生以技术服务为核心，凭借多年金融 IT 建设经验，以及对互联网的深刻洞察和理解，用优质的产品与服务，驱动金融机构创新发展。

### 【业务背景】

文本资讯是公司实体信息传播的主要媒介，精确定位发生新闻的公司实体直接决定如何开展下游金融工作。在金融资讯中，公司实体（公司实体数以千万计）多以领域简称的形式出现，从而引发歧义。例如，苹果是一个美国上市公司，也是一种水果。实体消歧旨在消除资讯处理过程中的实体多义问题，提纯文本信息。

歧义消除一般通过结合实体知识实现。近年来，人工智能技术飞速发展，使得众多问题的解决成为可能。我们希望能应用前沿人工智能方法，解决智能资讯中的实体歧义问题。

## 5. 任务要求

## 【开发说明】

本任务为金融领域中文文本的公司实体消歧系统构建。实体消歧(实体链接)的本质在于一个单词可能有多个意思,需通过上下文和知识库相关知识确定它所表达的确切含义。公司全称是确定性的,但在资讯、研报、问答等文本中,公司实体一般以简称的形式出现。同一个公司简称可能指向多个实体,例如,“恒生”可能指向恒生银行、恒生指数、恒生电子股份有限公司等。再比如,苹果是一个美国上市公司,也是一种水果,也是一种期货,还是大家熟知的手机 **iphone**。

实体消歧系统构建可包括知识库构建(维护实体知识,可包括目标实体的属性、文本描述等)、实体识别(识别命名实体,确定待消歧对象及其类别信息)和实体链接(与知识库中的对应实体进行关联)三个子任务。如有实体识别模块,实体识别模型的训练可使用额外的训练资源。

传统实体消歧任务主要是基于知识库的长文本,知识库较全,长文本具有更丰富的上下文信息以辅助实体消歧。本任务是基于垂直域(公司实体)消歧数据的实体消歧系统构建存在更大挑战:

- (1) 基于标注训练数据中生成知识库,且知识库需与下游消歧模型呼应;
- (2) 短文本上下文语境不丰富,知识库和语义理解需更加准确;
- (3) 与英文相比,中文由于自身的特点,短文本实体消歧更具挑战。

## 【技术要求与指标】

实体消歧系统输出结果包括文本中的特定命名实体及其消歧结果,模型评测通过对比模型输出和人工标注结果,计算准确率(Precision),召回率(Recall)和 F1 值(F1 score)。F1 值越高,模型排名越靠前。

给定文本输入(用 Query 表示),  $Q$  中  $N$  个实体 mentions、位置及其链接到是知识库的实体 id 人工标注结果为:  $ME_Q = \{(m_1, l_1, e_1), \dots, (m_k, l_k, e_k)\}$ 。相应地,对于模型输出结果为:  $ME'_Q = \{(m'_1, l'_1, e'_1), \dots, (m'_n, l'_n, e'_n)\}$ 。则有:

**精准率:**

$$P = \frac{\sum_{q \in Q} |ME_Q \cap ME'_Q|}{\sum_{q \in Q} |ME'_Q|},$$

**召回率:**

$$R = \frac{\sum_{q \in Q} |\mathbf{ME}_Q \cap \mathbf{ME}'_Q|}{\sum_{q \in Q} |\mathbf{ME}_Q|},$$

**F1 值：**

$$F1 = \frac{2 * P * R}{P + R}.$$

### 【任务清单】

待测试集（未标注，只有句子及编号信息）发布之后，参赛队伍提交此数据的消歧结果，以 `result.json` 命名（UTF-8 格式），系统会自动计算各队伍的评测结果。

测试结果数据规范为：

(1) 数据返回为 `dict`（字典形式）。`team_name` 声明队伍名；`submit_result` 为整个测试集对应结果的 `list` 串，每个元素对应一个测试集文本 `id` 的消歧结果。

(2) 每个测试集文本 `id` 对应的测试结果同样用 `dict`（字典）形式进行输出。包括，`text_id` 表示测试集 `id`；`text` 是与 `text_id` 对应的文本；`mention_result` 代表该测试文本的所有消歧结果构成 `list` 结构，可能有多个待消歧实体的消歧结果。

(3) 每个待消歧实体的消歧结果为一个 `dict`（字典）。包括，`mention`：实体提及（文本中出现的待消歧实体）；`offset`：`mention` 在句子中的起始位置（句首的位置为 0）；`kb_id`：消歧结果对应实体表中的编号（-1 表示 `mention` 指向并非我们所关系的实体）；`confidence`：置信度得分（不计入评分，只作为模型评测的辅助）。

(4) 提交结果的形式须参照上述数据结构，否则无法开展评测。

代码及方案文档需要打包成一个文件（`tar`, `zip`, `gzip`, `rar` 等均可），用 `code.xxx` 命名，要求提交所有的程序代码及相关的配置说明，程序应当可以运行且所得结果与 `result.txt` 相符。如果方法使用了额外资源，要求说明并提供资源文件或地址。

下面以两条测试数据为例，展示测试结果：

0 太阳能上半年净利润同比增长 3%；一块去获批挂牌新三板

1 各领域营销的能人智士一块去探讨、探索，共同努力

```
{
  "team_name": "ABC",
  "submit_result": [
    {
      "text_id": 0,
      "text": "太阳能上半年净利同比增长3%；一块去获批挂牌新三板",
      "mention_result": [
        {
          "mention": "太阳能",
          "offset": 0,
          "kb_id": 123,
          "confidence": 0.80685
        },
        {
          "mention": "一块去",
          "offset": 15,
          "kb_id": 254,
          "confidence": 0.783856
        }
      ]
    },
    {
      "text_id": 1,
      "text": "各领域营销的能人智士一块去探讨、探索，共同努力",
      "mention_result": [
        {
          "mention": "一块去",
          "offset": 10,
          "kb_id": -1,
          "confidence": 1
        }
      ]
    }
  ]
}
```

### 【提交材料】

- (1) 项目概要介绍；
- (2) 项目简介 PPT；
- (3) 项目详细方案；
- (4) 项目演示视频；
- (5) 团队自愿提交的其他补充材料。

### 【开发工具与数据接口】

本任务的标注数据包括训练集、开发集和测试集。前期将发布训练集和开发集，其中，知识库构建主要基于训练集，模型训练及选择基于训练集和开发集。测试集将在后期公布，用于对参赛队伍基于训练集和开发集构建的消歧系统进行评测。

训练集和开发集数据形式为，

text\_id: 测试数据集编号

text: 测试数据集编号对应的句子文本

mention\_result: 该句子的消歧结果 (list, 每个元素为一个 mention 的消歧结果)

mention: 实体提及 (待消歧的实体)

offset: 该实体提及在句子中的起始位置 (句首的位置为 0)

kb\_id: 消歧结果对应实体表中的编号 (-1 表示非领域词)

```
[
  {
    "text_id": 0,
    "text": "太阳能上半年净利同比增长3%；一块去获批挂牌新三板",
    "lab_result": [
      {
        "mention": "太阳能",
        "offset": 0,
        "kb_id": 123
      },
      {
        "mention": "一块去",
        "offset": 15,
        "kb_id": 254
      }
    ]
  },
  {
    "text_id": 1,
    "text": "隆重表彰为上海科技事业和现代化建设作出突出贡献的科技工作者",
    "lab_result": [
      {
        "mention": "上海科技",
        "offset": 5,
        "kb_id": -1
      }
    ]
  }
]
```

输入: text

输出: (mention, offset, kb\_id), 表示待消歧实体 mention (通过 offset 定位) 消歧之后在实体库中的 id

测试数据只提供可能包含待消歧实体出现的文本及其文本编号。例如,“各领域营销的能人智士一块去探讨、探索,共同努力”,该文本包含待消歧实体“一块去”,其对应编号 1; 文本编号 3 对应“食品股盘中大幅拉升,黑芝麻涨停”,其中“黑芝麻”需要消歧。具体形式如下:



- 0 太阳能上半年净利润同比增长5%；一块去获批挂牌新三板
- 1 各领域营销的能人智士一块去探讨、探索，共同努力
- 2 西安旅游发布年报，2018年，公司实现营业收入10亿元，同比增长10%
- 3 食品股盘中大幅拉升，黑芝麻涨停

每一行数据包括文本编号和文本，用\t进行分隔。测试过程中，对于单条测试文本，选手基于我们提供的待消歧实体列表对该文本中的待消歧实体进行消歧，消歧结果格式见测试结果提交见《任务清单》章节。待消歧实体列表形式如下：

平安银行	平安银行股份有限公司	000001
万科	万科企业股份有限公司	000002
国农科技	深圳中国农大科技股份有限公司	000004
世纪星源	深圳世纪星源股份有限公司	000005

分别对应待消歧实体、消歧后所关心的实体和消歧所后关心实体的编号。

## 6. 参考信息

无

## 7. 评分要点

本赛题评分要点参考附件一：A类企业命题统一评分标准。

除此之外，企业还会重点关注以下两个要素：

(1) 模型表现：F1值；

(2) 算法创新：算法独特、新颖，创新元素多，具有技术含量，可以有效地解决此类问题。

## 附件一：A 类企业命题初赛统一评分标准（仅供参考）

### 1、A 类企业命题初赛统一评分标准

内容		合计分值
项目创意	创意描述详细、清晰；对项目创意前景判断合理、准确；市场需求分析合理。创意独特、新颖，创新元素多，具有技术含量，有商业价值和社会应用价值。	20 分
实施方案	整体目标规划和工作进度安排合理；在各阶段工作目标清晰，难点明确，重点突出，解决方案合理并能兼顾目标与资源配置；操作周期和实施计划安排恰当。	30 分
技术实现与交付	技术路线清晰明确、技术工具成熟可靠；技术方案可行性高，项目完成度好；技术资源及经济成本控制合理，与项目需求匹配恰当。项目相关的知识产权证明（包括但不限于：专利证书、著作证书等）	30 分
项目展示	提交文档完整性、结构清晰合理、逻辑顺畅、文笔精炼。	20 分
总分		100 分