

“新国线杯”

第十五届全国大学生交通运输科技大赛

“交通大数据” 竞赛类竞赛方案

（第二部分 “物流大数据” 方向）

一、竞赛背景

网络货运平台兴起，如何通过大数据、物联网、人工智能等新兴技术，实现物流行业降本增效，一直是平台追求的目标。物流大数据主题竞赛旨在通过网络货运平台真实运单、订单、车型、位置等脱敏数据，利用工具软件和大数据挖掘与优化方法，在车货匹配、线路优化、客户需求分析等问题上寻求解决方案。

二、竞赛规则

1、数据来源

大赛数据为共生物流平台运营数据，共生物流平台为交通部“无车承运人试点企业”和工信部“大数据产业发展试点示范企业”，以先进技术建立 B2B2 车全链条物流互联网交易，提供物流运营、采购、销售、资金、财务、创业、人力资源共享等方面互联网支持服务，以共享提升物流服务和效率。

2、竞赛选题

在“交通大数据”主题竞赛类下，“物流大数据”竞赛方向围绕共生物流平台运营案例数据，设置车货匹配、车车联动、客户需求分析 3 个题目，参赛团队可任选其一，依据大赛组委会提供数据样本完成问题解决与作品设计。

参赛团队可使用各种优化工具进行问题的辅助解决，本次比赛主办方提供 Supply Chain Guru X 供应链优化工具（含运输优化模块）和 PMT 大数据挖掘工具，参赛团队可酌情使用。

2.1 题目一 车货匹配

2.1.1 业务说明

用户发布从 A 地发往 B 地物流运输需求，已知待承运货物体积（方）、重量（吨）、要求提货时间、要求到货时间、提卸地址经纬度等，根据当前时段周边车辆资源车型、当前位置，寻找最优匹配车辆。

2.1.2 数据说明

2.1.2.1 数据样例

表 1-1 车货匹配-运单数据

ID	运单编号	要求提货时间	要求到货时间	总方数	总吨位	总里程(km)
303307	20180402350177	2018-4-2 14:00:00	2018-4-2 16:00:00	30	7	140

表 1-2 车货匹配-地址数据

ID	运单ID	类型	省	市	区	地址	gpsX	gpsY	主节点	顺序
945126	348222	提货	广东	广州	天河区	广州市沙太南路342号	113.3348	23.17639	1	0

表 1-3 车货匹配-车型数据

ID	运单ID	车长(米)	车板
310905	303307	6.8	敞车
372403	348222	9.6	厢车

表 1-4 车辆基础数据

ID	车牌	车长(米)	车型	最大吨数	最大方数
24656	周 BX**X2	17.50	厢车	120.00	45.00

表 1-5 车辆 GPS 数据

vhc	gpsX	gpsY	gpsTime	province	city	area	addresses	Speed(km/h)
100990	120.10205200	30.09923000	2018-06-30T23:59:39.000Z	浙江省	杭州市	西湖区	[APP_IOS]G25长深高速	69.58799743652344

2.1.2.2 数据源说明

数据内容详见 “车货匹配-车型数据.txt” 、 “车货匹配-地址数据.txt” 、 “车货匹配-运单数据.txt” 、 “车辆基础数据.txt”、“位置基础数据.txt”。

2.1.3 解题要求

- (1) 车型匹配：按照待承运货物的重量和体积，推荐的匹配车型必须满足货物装载量；
- (2) 及时到车：根据提货地位置，检索周边在线车辆，确定车辆在要求提货时间点前 1 小时到车；
- (3) 返程优先：根据指定时段，判断车辆是否为返程车，返程车的成本一般较低，优先推荐；
- (4) 熟悉优先：车辆常跑线路，相对有价格优势，优先推荐。

2.1.4 结果内容

- (1) 综合考虑解题要求，设计车辆优先推荐模型。
- (2) 根据所设计的车辆优先推荐模型，输出 5 个按顺序优先推荐的车牌/ID（高->低）。

2.2 题目二 车车联动

2.2.1 业务说明

现有提货车辆在城市内多条线路装载零担货物，他们分别承担多个订单的提货任务，当前城市中目前已开设多个接驳网点，综合考虑车辆的行使线路、行驶里程、车辆的装载能力等，在完成全部提货任务的前提下，给出车辆最优汇合接驳网点的推荐顺序。

2.2.2 数据说明

2.2.2.1 数据样例

表 2-1 车车联动-订单数据表

ID	预约提货时间	提货省份	提货市	提货区	提货地址	重量(吨)	体积(方)	gpsX	gpsY
----	--------	------	-----	-----	------	-------	-------	------	------

1	1/26/2018 09:00:00	广东	广州	白云 区	庄中 路 XX 号	0.02	0.282	39.09316	43.739998
---	-----------------------	----	----	---------	-----------------	------	-------	----------	-----------

表 2-2 车车联动-提货车辆信息表

ID	车牌	车长（米）	车型	最大吨数	最大方数
25339	下 A6**F5	4.2	厢车	10	3

表 2-3 车车联动-城市接驳网点信息表

ID	网点名称	gpsX	gpsY	地址
1	接驳点 A	103.983361	30.506996	四川省成都市双留区腾飞十二路 971 号

2.2.2.2 数据源说明

数据内容详见“车车联动数据.xls”,包含订单数据,接驳网点,提货车辆数据三个子表。

2.2.3 解题要求

(1) 以全部提货车辆从同一接驳网点出发,完成相应提货任务并回到原出发接驳网点的总行驶里程最短为最优;

(2) 要求车辆在预约提货时间点前 20min 到车,装车时间按 20min 计算,车辆行驶速度按 60km/h 计算;

(3) 提货车辆的装载质量和体积不得超过其承重和体积限制。

2.2.4 结果内容

(1) 全部提货车辆最优汇合接驳网点的推荐顺序。

(2) 提货车辆选择及其行驶路线、行驶总里程。

2.3 题目三 客户需求分析

2.3.1 业务说明

客户过往的运单数据,可以分析出客户需求情况,根据过往的运单、承运车型,预估物流业务需求情况,分析出需求车型、线路分布,以及运量走势等。

通过总体的需求车型、线路,可以预估客户下一个月的日需求量(包括体积-总方数,重量-总吨位)。

2.3.2 数据说明

2.3.2.1 数据样例

表 3-1 客户需求分析-运单数据

ID	运单编号	要求提货时间	要求到货时间	总方数	总吨位	承运车辆
1	20180608907951	2019-12-13 00:00:00	2019-12-14 23:59:00	16	1.6	1

表 3-2 客户需求分析-地址数据

ID	运单 ID	类型	省	市	区	gpsX	gpsY	主节点	顺序
609626	247816	提货	陕西	咸阳	礼泉县	108.4826	34.59785	1	0
609627	247816	卸货	陕西	西安	雁塔区	109.0134	34.24416	1	0

2.3.2.2 数据源说明

提供 2018 年 1 月 1 日-2019 年 6 月 30 日的数据作为训练数据集； 2019 年 7 月 1 日-2019 年 7 月 31 日为需要预测的时间段。参赛者须基于此完成赛题。

数据内容详见“客户需求分析-地址数据（客户 A）.txt”、“客户需求分析-地址数据（客户 B）.txt”、“客户需求分析-运单数据（客户 A）.txt”、“客户需求分析-运单数据（客户 B）.txt”。

2.3.3 解题要求

（1）提供的数据为客户的持续 18 个月的发单体量。

（2）基于历史发单数据，考虑客户需求车型、总运量趋势、线路分布情况等维度进行每个客户的需求分布综合分析，其中线路分布以主提卸地址计算（如果多提多卸，以第一个提货地和最后一个卸货地计算）。

（3）预测每个客户在 2019 年 7 月份每天的运量（总方数和总吨位数）。

2.3.4 结果内容

- (1) 每个客户的需求分布综合分析。
- (2) 每个客户在 2019 年 7 月份每天的运量预测。

3、提交结果形式

为了便于评委了解作品及其技术细节，在作品提交截止日期前，各团队需要提供以下内容：

- (1) 模型文件或完整代码文件，结果输出文件；
- (2) 结题报告；

结题报告须对以下方面内容进行详细说明：

- (1) 研究背景：介绍研究的由来、意义、环境、状态、前人研究成果以及当前具有的条件等基础信息。
- (2) 研究数据支撑：介绍作品所使用到的比赛数据，并详细说明这些数据的使用情况。
- (3) 研究内容：对作品研究的内容进行梳理介绍。
- (4) 研究技术路线：详细介绍研究的方法、工具和手段，说明研究的流程。
- (5) 研究结果：对研究的结果进行完整、详细、清晰表达。
- (6) 研究总结：对研究的问题、过程及成果进行总结。

4、评比规则

本次大赛坚持公开、公平、公正的原则对所有符合要求的参赛作品进行评审，评审内容包括：

- (1) 作品报告完整性（分数占比 10%）
- (2) 研究成果的合理性、准确性和实用性（分数占比 45%）
- (3) 研究方法的科学性、先进性和创新性（分数占比 45%）

三、竞赛组织

“物流大数据”竞赛方向属于“交通大数据”主题竞赛类的方向之一，其参赛要求、组织程序和评审办法参照《“新国线杯”第十五届全国大学生交通运输科技大赛实施方案》执行，优秀作品将有机会入围“交通大数据”主题竞赛类决赛答辩。“交通大数据”主题竞赛类，每个高校推荐作品不得超过 4 件。

四、其他说明

1、Supply Chain Guru X 供应链优化工具（含运输优化模块）和 PMT 大数据挖掘工具的基本介绍见附件一和附件二。

更多相关工具问题咨询,请加入 QQ 群:978903563 或咨询电话:18001228890, 联系人: 阎海伟。

2、工具和数据下载平台目前采用申报方式开放注册。有意参加“物流大数据竞赛方向的高校请填写下表发送至组委会邮箱 nactrans2020@bjtu.edu.cn (邮件名称格式:“物流大数据”账号开通申请-高校名称), 向组委会提出账号开通申请。

原则上每个学校参赛队伍不超过 5 组, 若有需要注册更多队伍请与组委会说明情况。

表 1: 注册报名表样式

队伍编号	成员姓名	学校名称	指导老师	联系人姓名及电话
1	成员 1(队长)			成员 1 (12345678910)
	成员 2			
	...			
2				

各参赛队伍报名成功后, 可从以下地址获取相关工具和数据:

<http://contest.cfnet.org.cn/logistics2020/index.html>

3、未尽事宜, 解释权及修改权归大赛组委会。

附件一、Supply Chain Guru X 供应链设计工具介绍

1、工具简介

Supply Chain Guru X (SCGX) 供应链设计工具可以系统化建立“数据驱动”的供应链模型，并通过云计算、数据集成管理、模拟仿真、数据可视化等技术手段，诊断和分析供应链存在的问题，并基于优化目标输出优化结果，解决供应链网络优化、库存优化、产销协同、运输优化等方面的问题。

2、工具说明

Supply Chain Guru X 集成网络优化、库存优化、运输优化和模拟仿真四大优化引擎。其中，利用其运输优化模块，能够在综合考虑各种成本、时间限制、车辆资源、交通限制、运输资源的条件下，进行多点路径优化。

可为以下经典问题设计解决方案：

- (1) 取货运输任务整合
- (2) 配送运输任务整合
- (3) 周期运输任务平衡及排车计划
- (4) 回程优化
- (5) 配送及取货任务交错的线路
- (6) 车辆数量优化
- (7) 通过中转场地优化运输任务



图 1 多站点运输线路优化

3、建模过程

运输优化建模包含 5 个步骤：

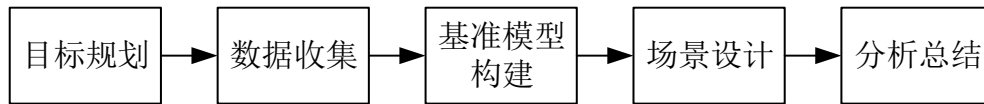


图 2 运输优化建模流程

(1) 目标规划：包含运输优化问题的分析及目标设定、运输优化模型假设、运输优化场景设定。

(2) 数据收集：包括数据分析维度确定、数据获取与数据清洗、数据导入与映射关系。

(3) 基准模型构建：构建并验证基准模型。

(4) 场景设计与分布：包括应用场景设计、建模与分析。

(5) 分析总结：分析并总结问题，给出建议决策。

附件二、PMT 大数据挖掘工具

1、工具简介

PMT 大数据挖掘工具采用可视化建模的设计思路（即用图形化的方法，来建立整个挖掘流程），支持数据预处理、数据探索性分析、建立模型、模型评估、模型应用整个数据挖掘过程，在建立模型的核心模块，提供解决分类预测、回归预测、聚类分析、关联规则、序列分析、异常检测等数据挖掘问题的主流算法，并且具有可拓展性功能，通过编写 Python 脚本的方式，实现自定义分析功能。

2、工具说明

PMT 对数据按照特定的方式方法进行分析处理，发现隐藏在数据背后的问题和规律，包含从原始数据的预处理、加工、挖掘、分析到发现规律，最后再到结果可视化呈现的全过程。具备数据描述与理解功能、数据获取与组织功能、数据挖掘建模功能。

封装了包括机器学习、数据预处理和数据可视化等组件，目标是以一种最为简约的方式来解决具体业务场景中的问题，所提供的核心功能如下：

(1) 数据预处理：数据记录选择、数据特征选择、数据标准化、

缺失值处理、PCA 降维、数据合并、离散化、异常值处理等；

(2) 数据可视化：直方图、箱线图、树图、散点图、维恩图、地图、序列图、相关图等；

(3) 数据建模：多元回归、决策树、逻辑回归、神经网络、ARIMA、随机森林、支持向量机、关联规则、聚类算法等；

(4) 模型评估：混淆矩阵、ROC 曲线、误差分析 (RMSE, MSE, MAE) 等。

3、建模过程

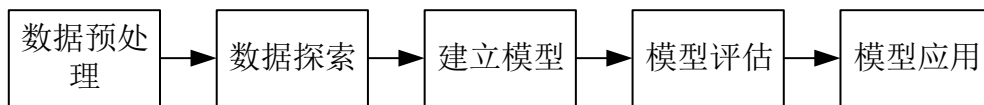


图 3 数据挖掘分析流程

(1) 数据预处理：数据预处理的目的是让数据适应模型，匹配模型的需求，包括数据转换，数据集成等处理内容；

(2) 数据探索：数据探索是对建模分析数据进行先导性的洞察分析，利用绘制图表、计算某些特征量等手段，对样本数据集的结构特征和分布特性进行分析的过程；

(3) 建立模型：数据挖掘的核心阶段是基于既定的数据和分析目标选择适宜的算法模型进行建模训练和迭代优化；

(4) 模型评估：模型评估是评估所构建的模型是否符合既定的业务目标，它有助于发现表达数据的最佳模型和所选模式将来工作的性能如何。模型评估秉承的准则是在满足业务分析目标的前提下优先选择简单化的模型；

(5) 模型应用：模型部署及应用是将数据挖掘结果作用于业务过程，即将训练得到的最优模型部署到实际应用中。