

个人简介

基本情况

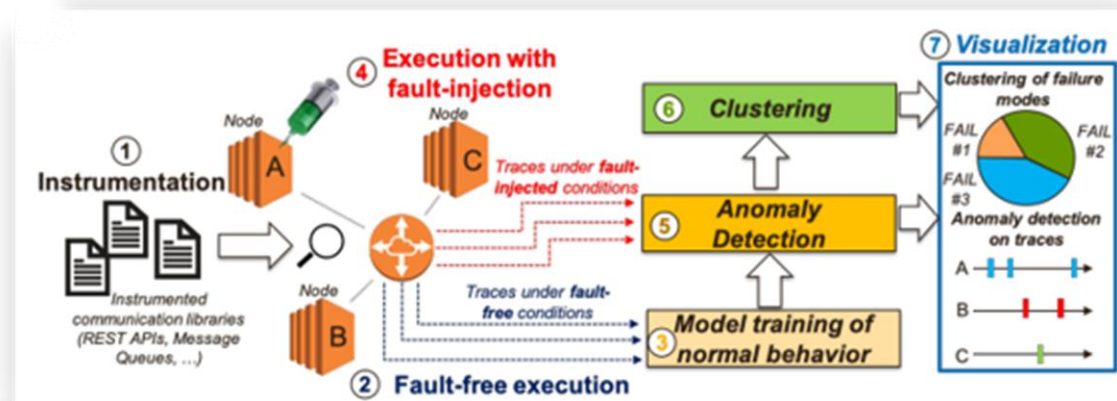
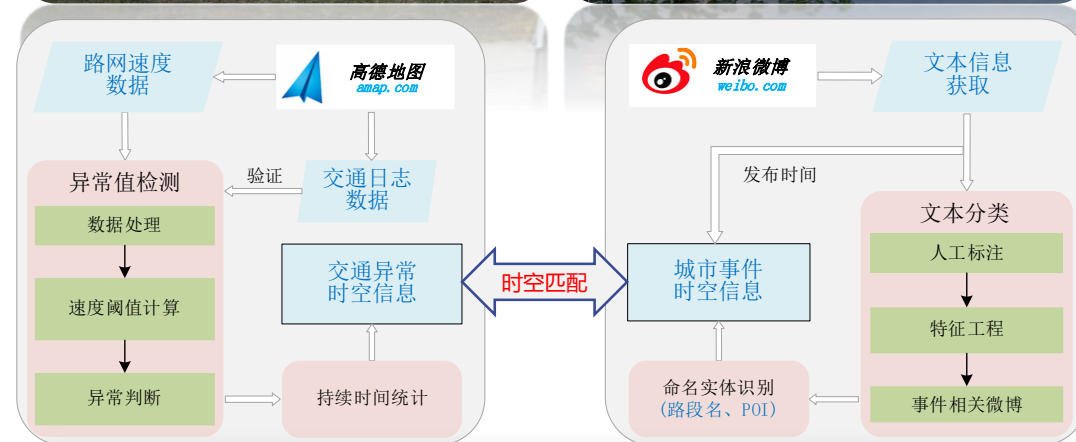
- 于晓飞，北京航空航天大学交通运输工程硕士研究生
- 个人主页: <https://xiaoffy.netlify.app/>

科研方向

- ✓ 基于数据驱动的选址建模优化问题研究
- ✓ 城市路网异常态势感知与预测
 - **异常检测与解析**: 交通异常检测、城市事件挖掘
 - **交通态势演化**: 路网状态预测、拥堵传播机理

兴趣领域

- **KPI异常检测**，根因诊断方法分析；
- 基于机器学习算法在异常检测方面可解释性研究；
- 数据挖掘及风险控制等算法研究；



竞赛获奖

论文

- Evaluating and Predicting Road Network Resilience Using Traffic Speed and Log Data [C] 22nd CICTP, 2021.(Best Paper Award)
- 基于数据驱动的停车设施选址研究（交通信息与安全）

专利

- 城市交通状态与安全态势协同关联分析平台
- 一种考虑长短期路段性能的道路韧性衡量方法
- 一种基于数据驱动的停车设施选址方法

竞赛

- 华为杯数学建模竞赛 全国三等奖2021
- 大连市优秀毕业生2020
- 国家级创新创业项目2019
- 全国大学生数学竞赛 全国三等奖2018
- 全国大学生数学建模竞赛 辽宁省二等奖2018
- 全国大学生数学竞赛 辽宁省一等奖2018
- 国家级创新创业项目
- CICPT 2022 国际学术会议最佳论文奖

学术兼职

- IEEE TITS 审稿人

科研项目

一. 交通大数据分析

二. 路网韧性研究

三. 公交线网优化

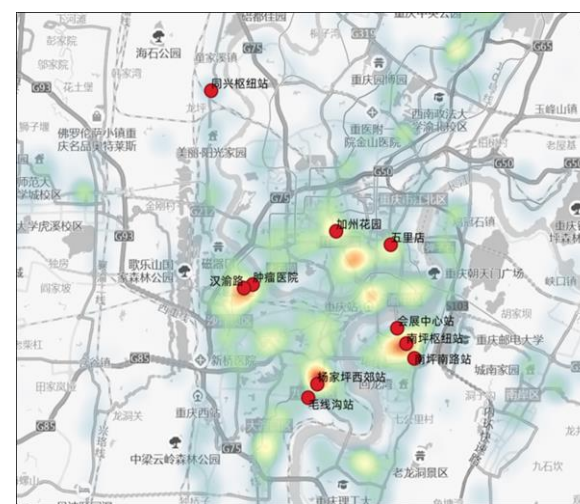
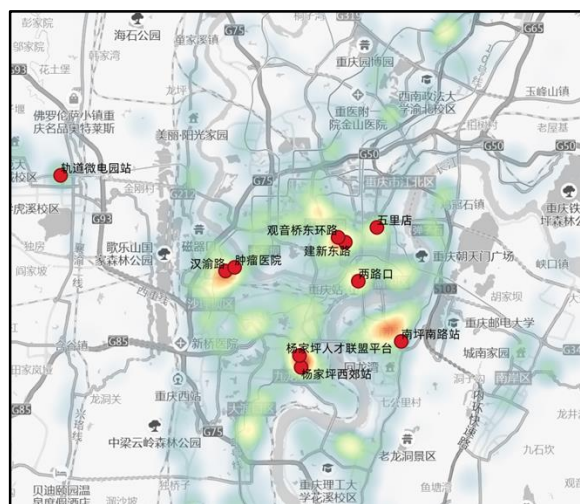
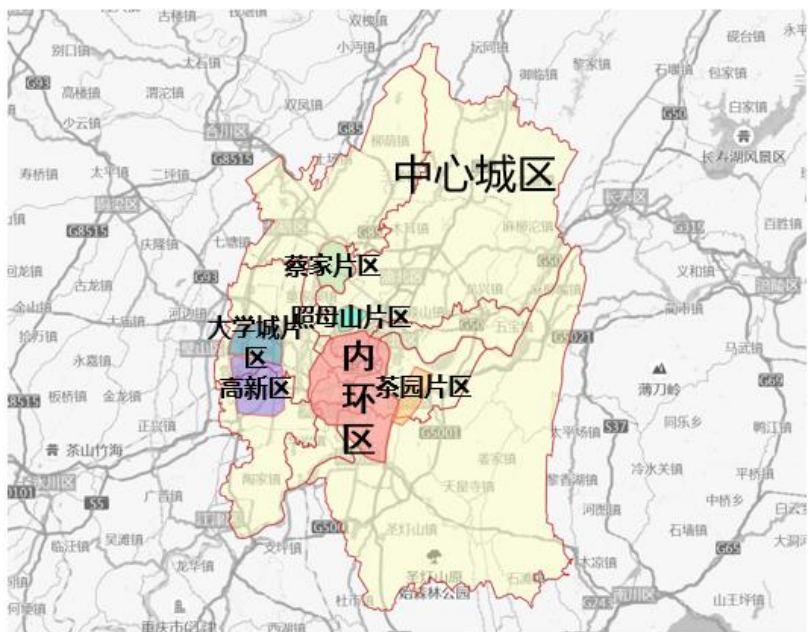
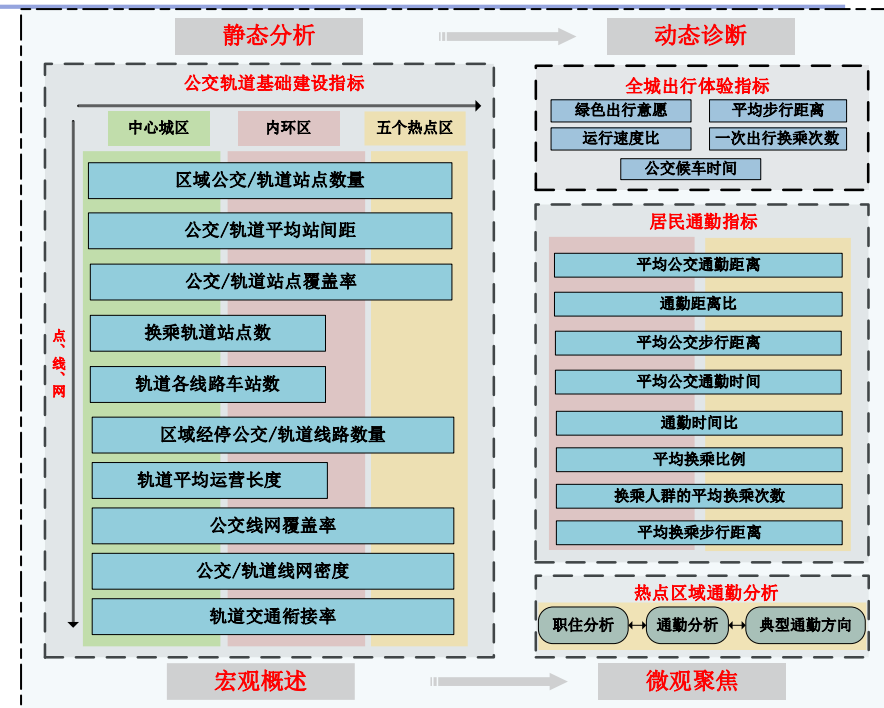
四. 平台开发

五. 数据挖掘—风控

六. 实习经历

1.重庆市主城区公共交通动态判断与对策研究

- 作为学生主力，与高德对接，梳理**静态指标**，**出行体验指标**和**通勤指标**三大部分指标计算文档与报告大纲
- 驻场高德地图和重庆城市交通投资开发集团，利用python和QGIS，处理并可视化线路基础数据和海量通勤数据
- 从公共交通基础设施建设，居民出行和通勤分析三个方面，按照重庆中心城区，内环区和五个热点区三个层级展开分析
- 主导五个热点区通勤思路分析，挖掘通勤需求与公共交通供给间的关系，给出优化建议，部分建议已经落地实施



2.交通韧性（国家重点研发计划）

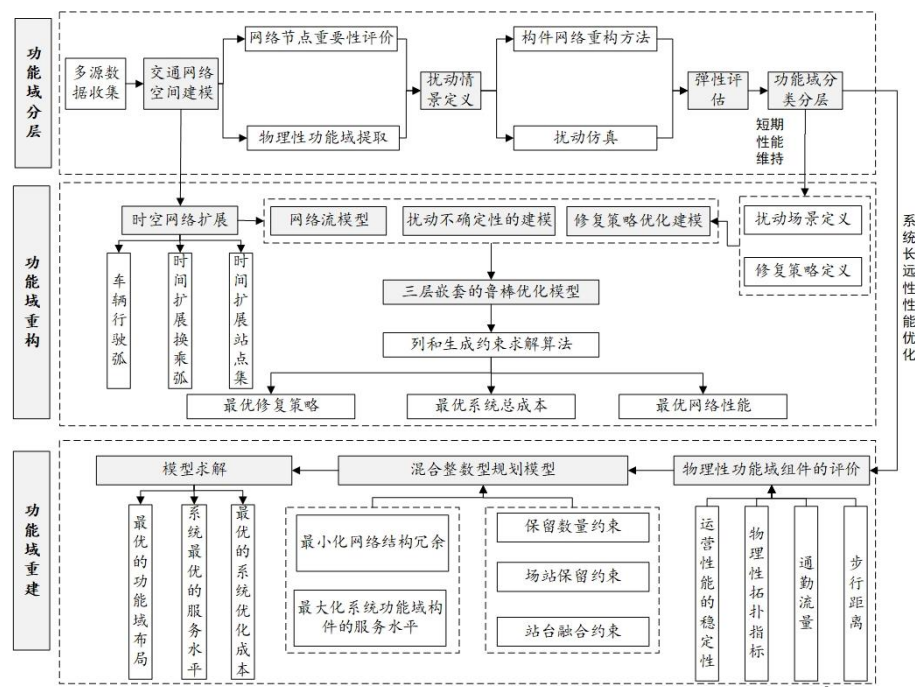
Evaluating and Predicting Road Network Resilience Using Traffic Speed and Log Data

- 研究交通弹性和速度之间的关系以及在路网中的传播机理
- 以北京市朝阳区部分区域为例进行验证，基于车辆速度数据
- 根据弹性在路网中的传播，更好的对路况交通进行预测预警



一种考虑长短期路段性能的道路韧性衡量方法

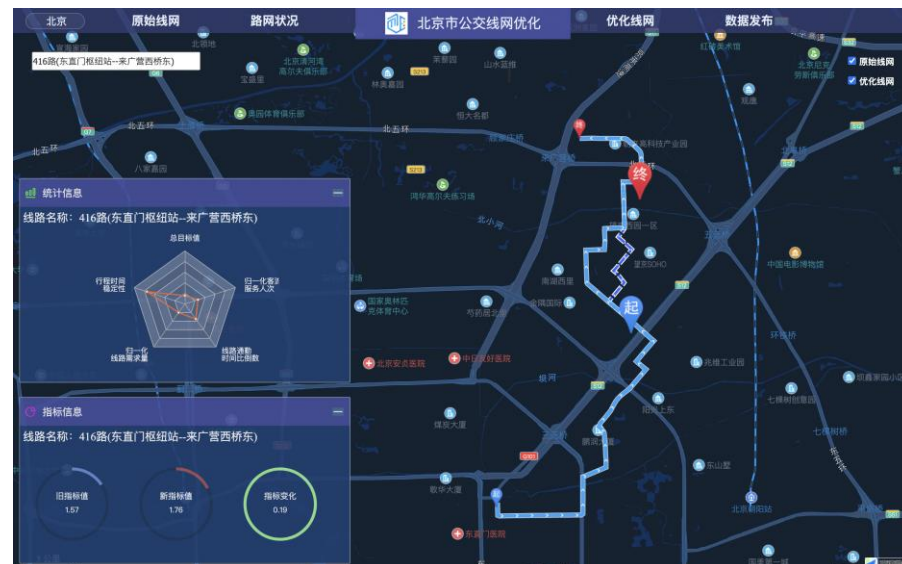
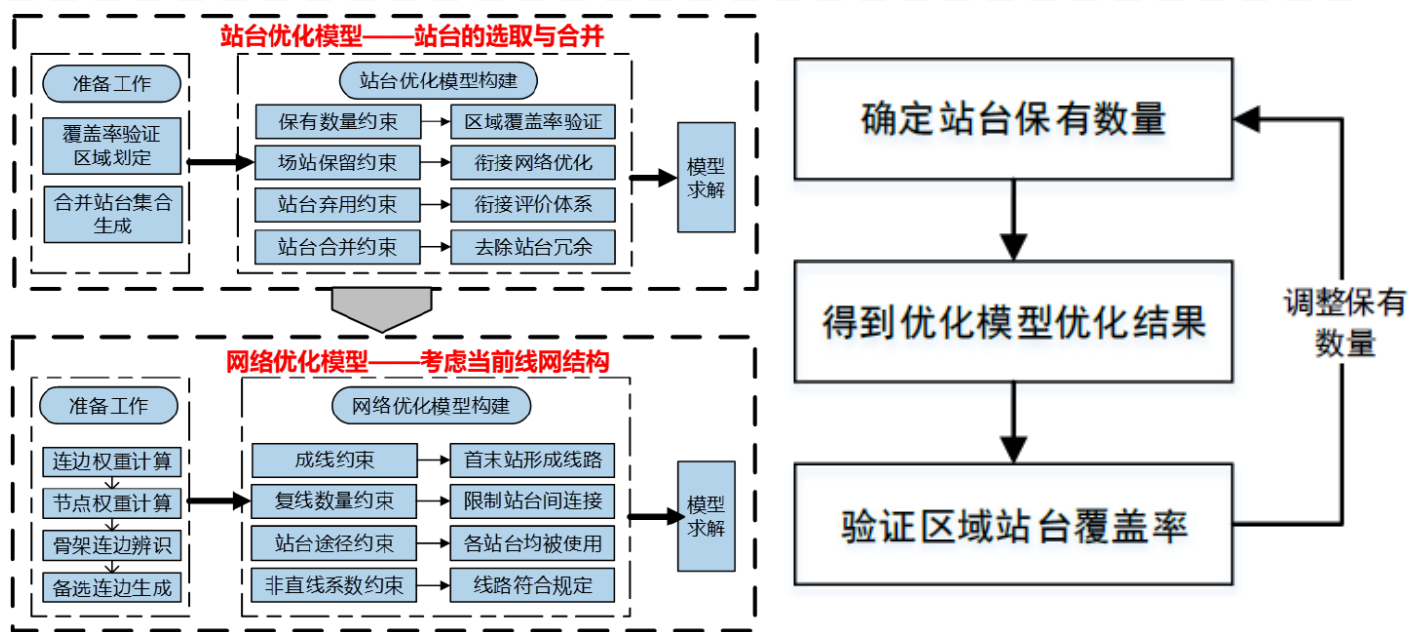
- 基于数据驱动同时考虑道路长期和短期性能实现道路韧性评估
- 收集并处理路网速度数据，统一不同等级道路的度量方法
- 通过计算路段速度方差确定波动较大的区域作为最终的韧性测量区域
- 基于历史数据计算各个路段的自由流速度，计算考虑长短期性能的路段韧性并进行灵敏度分析。



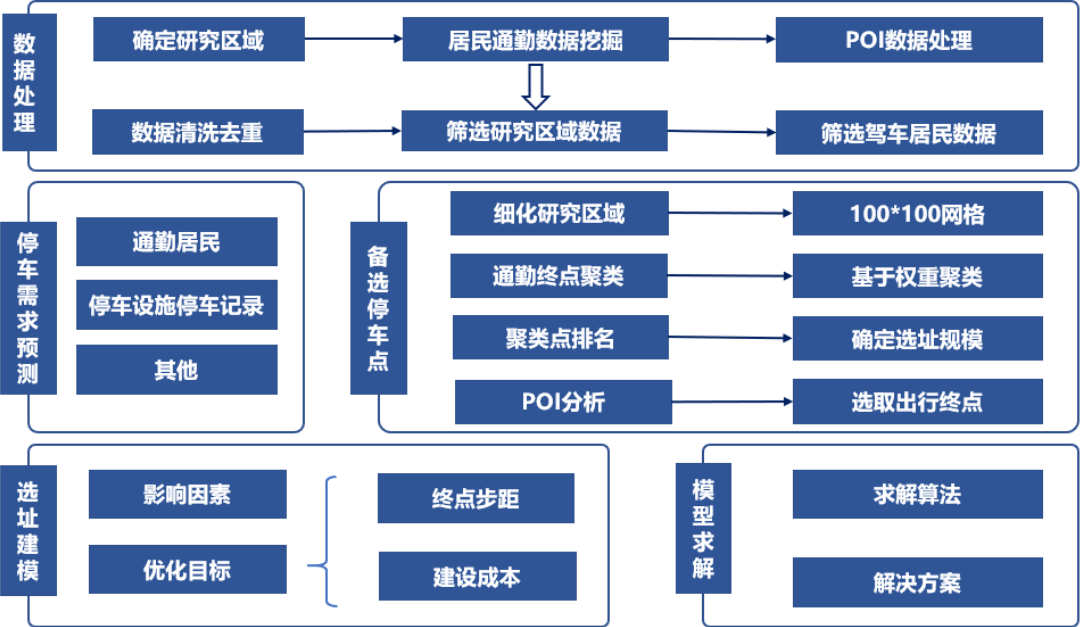
3.公交线网优化

公交线网优化

- 基于互联网定位数据对居民职住信息及通勤方式进行挖掘。
- 提出了一套基于互联网定位数据考虑居民出行体验的公交线网评价指标和基于互联网定位数据与评价结果考虑当前线网结构的公交线网优化流程。
- 辅助公交集团在保持整体线网稳定的情况下实施大规模线网优化，减少站台冗余，提升居民出行体验与公交竞争力。



4.停车设施选址研究



➤ 停车设施选址模型—基础模型

符号	含义
M	备选停放点集合
N	停车需求点集合
U_{min}	停放点最大容量
U_{max}	停放点最小容量
u_j	第j个需求点的停车需求量
d_{ij}	需求点j到停车设施点i的步距
x_i	0-1变量, 表示是否选择第i个备选停放点
f_i	第i个停放点的固定建设成本
q_i	表示第i个停放点的维护成本
s_i	表示第i个停放点的容量
y_{ij}	0-1变量, 表示第j个需求点的车辆是否停在第i个备选停车点
P_{ij}	第j个需求点的车辆停在第i个停车点的车辆

$$\min(\sum_{i=1}^m (f_i x_i + q_i s_i) + \sum_{i=1}^m \sum_{j=1}^n d_{ij} y_{ij} p_{ij})$$

s. t.

- (1) 容量约束: $\sum_{j=1}^n y_{ij} p_{ij} \leq s_i$
- (2) 设施服务约束: $\sum_{i=1}^m y_{ij} = 1, \forall j \in N$
- (3) 供需约束: $x_i \geq y_{ij}, \forall j \in N, \forall i \in M$
- (4) 供给松弛约束: $\sum_{i=1}^m p_{ij} y_{ij} \geq u_j, \forall j \in N, \forall i \in M$
- (5) 容量边界约束: $U_{min} \leq s_i \leq U_{max}, \forall i \in M$
- (6) 假设条件: $p_{ij} = \begin{cases} u_j & \text{if } y_{ij} = 1, \forall j \in N, \forall i \in M \\ 0 & \text{if } y_{ij} = 0 \end{cases}$
- (7) 0-1约束: $x_i \in [0, 1], \forall i \in M$
- (8) 0-1约束: $y_{ij} \in [0, 1], \forall j \in N, \forall i \in M$
- (9) 单一容量约束: $p_{ij} \leq s_i, \forall j \in N, \forall i \in M$
- (10) 单一需求约束: $p_{ij} \leq u_j, \forall j \in N, \forall i \in M$

```
m1.addConstrs(((sum(p[i,j] for j in range(n))<=s[i]) for i in range(m)),name='v1')
m1.addConstrs((p[i,j]+x[i]*M>=0 for i in range(m) for j in range(n)),name='v1-1')
m1.addConstrs((p[i,j]-x[i]*M<=0 for i in range(m) for j in range(n)),name='v1-2')
m1.addConstrs((p[i,j]>=0 for i in range(m) for j in range(n)),name='v1-3')
m1.addConstrs((sum(y[i,j] for i in range(m))==1 for j in range(n)),name='v2')
m1.addConstrs(((x[i]>=y[i,j]) for i in range(m) for j in range(n)),name='v3')
m1.addConstrs((sum(p[i,j]*y[i,j] for i in range(m))>=ss[j] for j in range(n)),name='v4')
m1.addConstrs((s[i]>=Umin for i in range(m)),name='v5.1')
m1.addConstrs((s[i]<=Umax for i in range(m)),name='v5.2')
```



对需求不确定性模型进行求解分析, 通过hoeffding不等式进行安全近似约束推导验证模型正确性, 提出选址方案。7

5.机组排班问题—优化

- 采用线性加权的方法对多目标的目标函数进行处理，构建混合整数规划模型来精确描述问题，并设计启发式算法来对模型进行求解，最终编程实现算法完成了给定算例的求解和结果的分析。
- 创新性地提出了基于 LS 的“解编和组编”优化思路，具有较强的适应性和推广性；构建的算法能在短时间内取得较为理想的解方案，实用性较强。

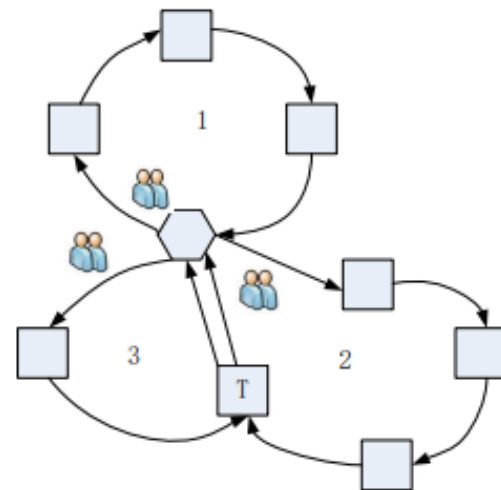


图 3.4 初始解构建示意图

表 3.2 基于 LS 的优化算法

```
1: 输入初始解  $S_0$ 
2: 初始化外部循环迭代次数  $outNum$ ，外部循环变量  $i$ ，内部循环迭代次数  $inNum$ ，内部循环变量  $j$ ，全局最优解  $bestSol \leftarrow S_0$ ，局部最优解  $locSol \leftarrow S_0$ ，次优解  $worSol \leftarrow S_0$ 
3: while  $i++ < outNum$ , do
4:    $j = 0$ 
5:   while  $j++ < inNum$ , do
6:     随机选择  $locSol$  中的任意两个航班环 A、B
7:     if A、B 有公共机场 T, then
8:       以 T 为机组人员重组点对 A、B 进行一次组编和解编优化操作
9:     else if A、B 有公共基地 D, then
10:      以 D 为机组人员重组点对 A、B 进行一次组编和解编优化操作
11:     else
12:      重新选择用于优化的航班环 A、B
13:   end while
14:   end if
15:   if  $f(locSol) < f(bestSol)$ , then
16:     更新  $bestSol \leftarrow locSol$ 
17:   end if
18:   if 达到模拟退火接受标准, then
19:      $locSol \leftarrow worSol$ 
20:   else
21:      $locSol \leftarrow bestSol$ 
22:   end if
23: end while
24: 更新模拟退火参数相关参数
25: end while
26: 输出最优解的目标函数值  $f(bestSol)$ 
```

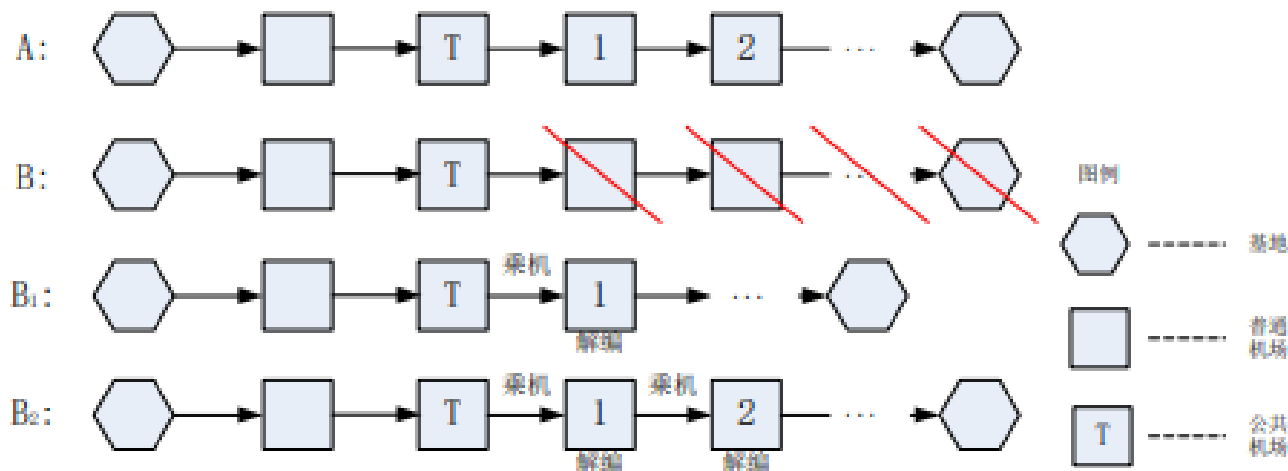
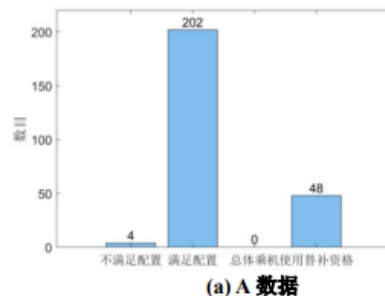
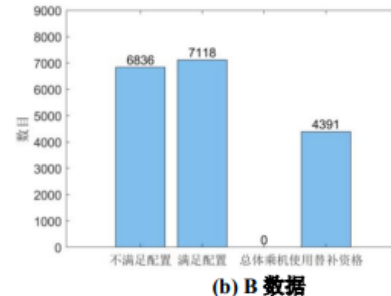


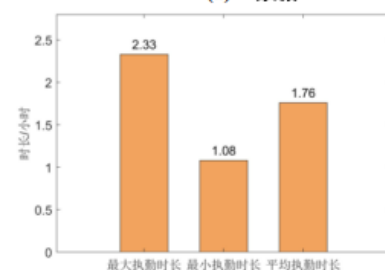
图 3.4 有公共机场 T



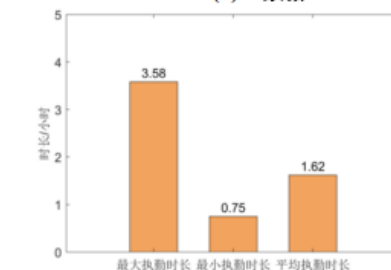
(a) A 数据



(b) B 数据



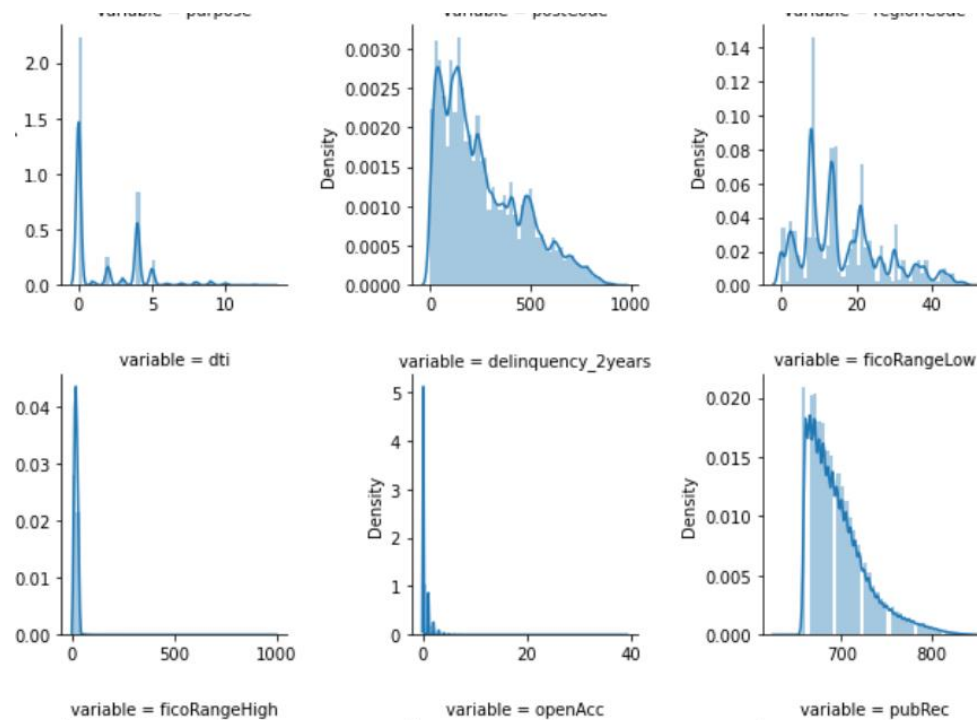
(c) A 数据



(d) B 数据

6. 数据挖掘算法相关

- 根据信贷平台的贷款数据，47个特征变量以及15个匿名变量对借贷人违约的可能进行预测。
- 本项目的技术路线以及算法设计流程：
 - 1、数据探索阶段理清各数据特征之间的关联性，对变量类型进行区分性考虑；
 - 2、结合已有金融知识的了解进行特征工程的构建，探索性的构建衍生特征，并对连续性特征进行分箱处理。对缺失值，异常值 进行处理，根据PSI指标进行特征提取；
 - 3、构建 XGBoost, LGBM, GBDT 等模型，通过 网格调参 获得最佳参数值，通过AUC等指标进行结果的评估。探索性尝试 CatBoost 建模并取得较好效果，采用采样的方法对不平衡样本进行探索性处理。

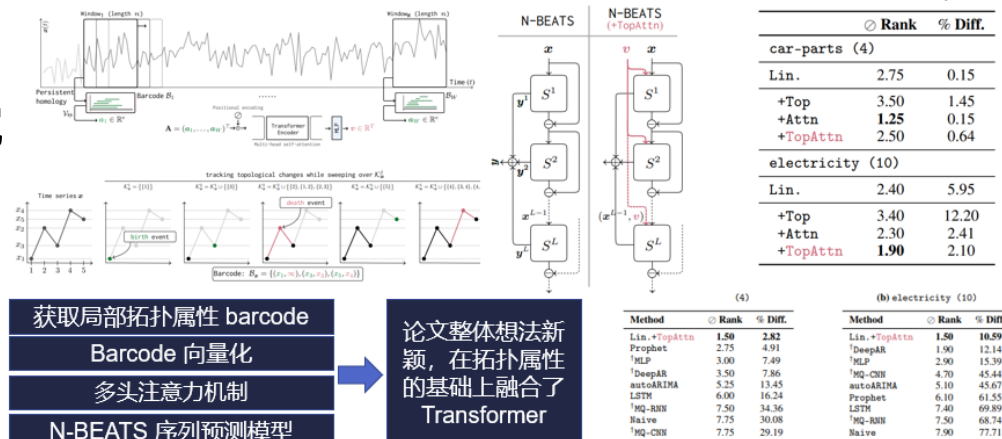


delinquency_2years	借款人过去2年信用档案中逾期30天以上的违约事件数
ficoRangeLow	借款人在贷款发放时的fico所属的下限范围
ficoRangeHigh	借款人在贷款发放时的fico所属的上限范围
openAcc	借款人信用档案中未结信用额度的数量
pubRec	贬损公共记录的数量
pubRecBankruptcies	公开记录清除的数量
revolBal	信贷周转余额合计
revolUtil	循环额度利用率，或借款人使用的相对于所有可用循环信贷的信贷金额

7.联想集团：智能运维算法研究员

- 追踪智能运维领域前沿研究，对关键性指标的异常检测、时序预测、根因分析等方向算法进行梳理，在组内做学术分享。
- 基于CPU,内存等历史数据和实时数据进行异常检测算法的研究，探索无监督学习算法在异常检测方向的应用及可解释性。
- 基于AIOps 2022国际运维挑战赛数据，从服务器节点角度出发进行故障识别与分类，指标之间关联关系的确定以及完成对多维KPI异常定位的工作，通过注入故障来验证算法的效果，探索AutoEncoder和VAE算法在样本不均衡条件下的效果。

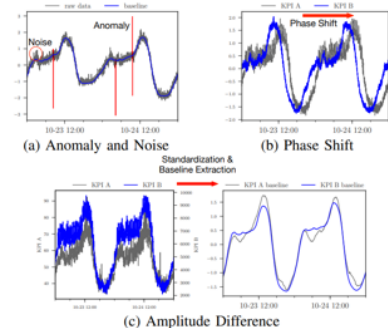
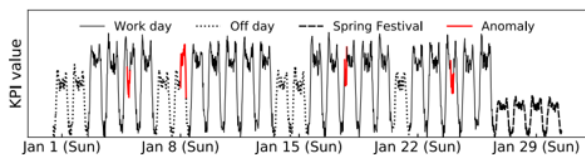
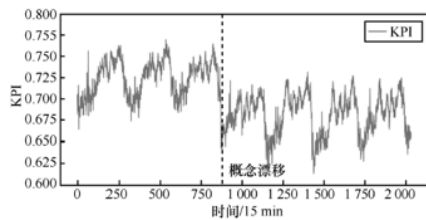
针对时间序列预测的拓扑注意力模型 TopAttn & N-BEATS



TOPOLOGICAL ATTENTION FOR TIME SERIES FORECASTING, NIPS 2021

节假日效应
受节假日影响，KPI发生异常波动，模型通用化，简单调节参数是不可取的。

概念漂移
KPI 序列受活动发布、业务变更、资源调度等特殊事件影响产生的水平漂移，预期漂移不属于异常，意外漂移属于异常

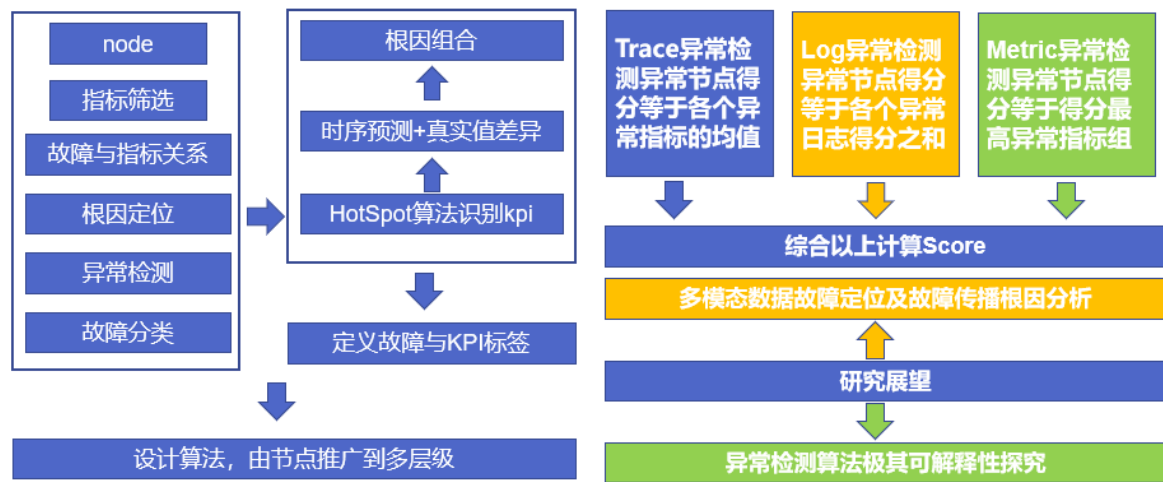


海量KPI异常检测

基于时间序列聚类的方法，根据每个KPI的基本形状基于DBSCAN等算法生成簇，进行异常检测。

- Automatic and Generic Periodicity Adaptation for KPI Anomaly Detection
- Robust and Rapid Clustering of KPIs for Large-Scale Anomaly Detection

微服务架构电商系统下故障识别与分类



8.中国工商银行总行：反欺诈算法研究

- 1.专注于数据挖掘，金融风控在反欺诈，反电诈方面的研究，并对已有算法进行调研和探索。
- 2.对团伙欺诈进行识别，研究图算法等社区发现算法进行业务的基本应用。

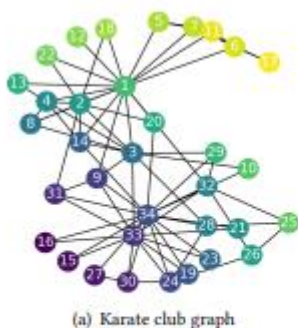
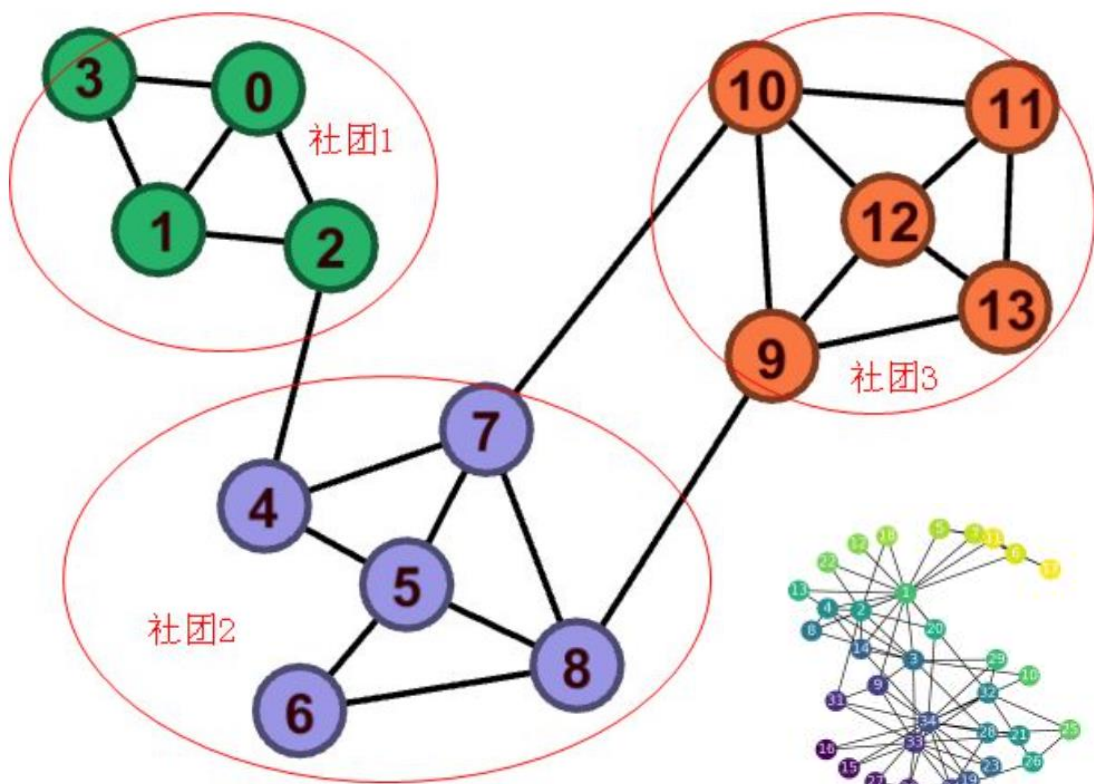
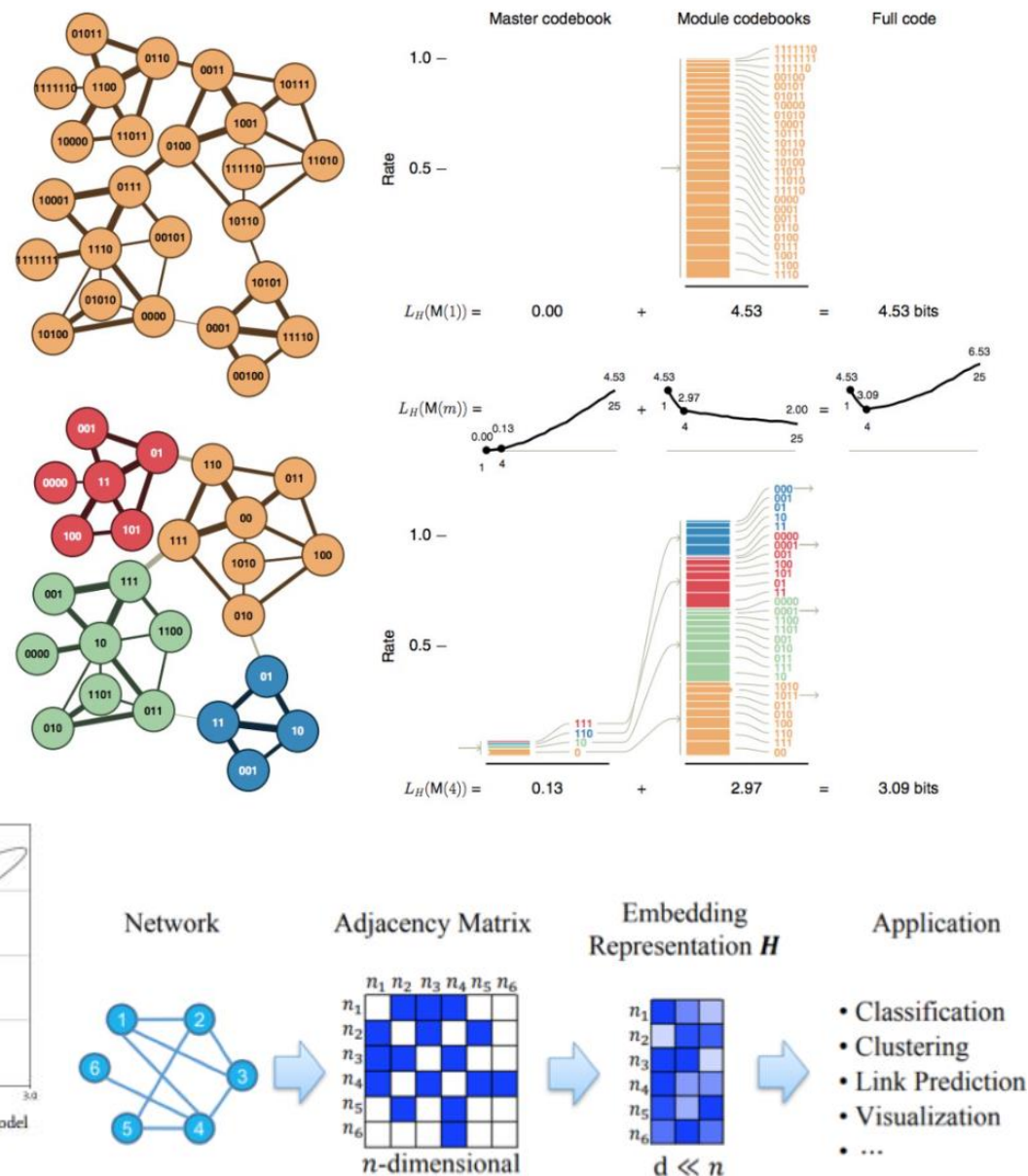


Figure 1: Embedding nodes and communities in a 2D space.



9.求职意愿

能做什么

- 数据分析相关工作
- 机器学习/深度学习
- 运筹优化算法
- 前端开发/Vue
- 科研探索 (Research)
- 程序设计类工作

想做什么

- 数据挖掘/时序预测/根因分析/异常检测/金融风控
- 智慧城市/智能交通相关场景算法研究
- 推荐/搜索相关算法研究

Q & A