

基于行车轨迹数据估计交通信号灯周期

摘要

地图导航现在已经成为日常交通出行中必不可少的一环，其中精确的红绿灯数据可以帮助司机和行人来更好的进行道路选择和驾驶决策。现有地图厂商希望获取城市路网中所有交通信号灯的红绿周期，并计划使用大量客户的行车轨迹数据来估计交通信号灯的周期。

针对问题一，我们整理了**车辆轨迹数据序列**进行，并对**车辆道路行驶行为**进行了定义和判定，进而计算出每个交通路口的**减速制动距离**。同时建立了**交通路口简化模型**，定义了**红绿灯决策区域**，并根据决策区域中的车辆道路行驶行为获取了**红绿灯特征值的离散时间序列**。最后，使用**遗传算法**寻找时间序列中的理想转变点，最终求得**固定红绿灯周期**。

针对问题二，由于数据集中仅为部分**样本车辆**的行车轨迹，并且轨迹数据存在定位误差，我们建立了**误差灵敏度分析模型**，对**车流量**、**样本占比**和**定位误差**进行了量化，并基于量化结果来计算**有效信息的损失量**。最后，我们基于误差分析对给出的数据集数据进行**补充和优化**并求取了红绿灯周期。

针对问题三，面对红绿灯周期可能发生变化的情况，我们使用**密度抽样**对变化进行检测，根据抽样异常点对**数据分段**，并对分段数据计算**变化衡量因子**来确定异常点是否为所求红绿灯周期转变时间序列点。按照上述流程对检测序列进行逐一检查来确定最终的红绿灯周期转变点并求取各阶段的红绿灯周期。

针对问题四，我们对路口四个方向的数据集进行了**数据可视化**，并以此建立了**四方向交通路口模型**。进一步，我们求取**四方向车辆位置时间序列**，并转化为最终的离散红绿灯特征值时间序列，最终使用问题一中的数学模型求解最终的红绿灯周期。

关键词：时间序列, 车辆轨迹, 红绿灯周期, 遗传算法, 车辆道路模型, 特征值

目录

1	问题重述	3
2	模型假设	3
3	符号说明	4
4	模型的建立与求解	5
4.1	问题一模型的建立与求解	5
4.1.1	车辆轨迹数据序列	5
4.1.2	车辆道路行为	5
4.1.3	车辆直行运动状态	6
4.1.4	减速制动距离	7
4.1.5	交通路口简化模型	7
4.1.6	红绿灯决策区域	8
4.1.7	红绿灯离散时间序列	9
4.1.8	确定理想时间序列点算法	12
4.1.9	问题一模型的求解	13
4.2	问题二模型的建立与求解	14
4.2.1	误差灵敏度分析模型	14
4.2.2	基于误差分析的数据优化	15
4.2.3	问题二模型的求解	15
4.3	问题三模型的建立与求解	16
4.3.1	变化检测密度抽样	16
4.3.2	分段数据的变化检测	16
4.3.3	遍历确定周期变化点	17
4.4	问题三模型的求解	17
4.5	问题四模型的建立与求解	18
4.5.1	四方向交通路口模型	18
4.5.2	四方向车辆位置时间序列	19
4.5.3	问题四模型的求解	19

1 问题重述

某电子地图服务商希望利用车辆行车轨迹数据估计城市路网中所有交通信号灯的红绿周期，以提供更好的导航服务。由于许多信号灯未接入网络，无法直接获取其周期信息，也不可能所有路口安排人工读取。因此，该公司计划利用大量客户的行车轨迹数据估计信号灯的周期。以下是要完成的任务：

1. 若信号灯周期固定不变，且已知所有车辆的行车轨迹，建立模型，利用车辆行车轨迹数据估计信号灯的红绿周期。
2. 考虑到只有部分用户使用该公司的产品，即只能获取部分样本车辆的行车轨迹，并且轨迹数据存在定位误差，讨论样本车辆比例、车流量、定位误差等因素对上述模型估计精度的影响。
3. 检测信号灯周期是否有可能发生变化，以及如何尽快检测出这种变化，并获取变化后的新周期。
4. 识别出某路口连续 2 小时内所有方向样本车辆的信号灯周期。

2 模型假设

1. 假设车辆在直行时的运动状态只有匀速、加速和减速。
2. 假设路口一段时间内的车流量在总体上保持平稳。
3. 假设定位带来的误差符合正态分布。
4. 假设司机在同样的交通条件下会作出同样的道路驾驶决策。
5. 假设路口类型只有直行路口、丁字路口和十字路口三种。

3 符号说明

符号	说明
T_d	ID 为 d 的汽车轨迹段
DA	车辆道路行为
$\bar{v}$	采样点的平均速度
H_{down}	减速制动距离
S_{rg}	红绿灯总周期
T_{sum}	红绿灯总周期
$State$	红绿灯特征值
TV	车流量
λ	样本车辆比例
$Loss$	有效信息损失量
C	变化衡量因子

4 模型的建立与求解

4.1 问题一模型的建立与求解

4.1.1 车辆轨迹数据序列

对于驶入和驶出路口每一车辆，其车辆行驶过程中的轨迹数据特征由车辆 ID、当前位置 X 坐标、当前位置 Y 坐标和时间戳 4 种信息构成，关注车辆行驶路径对象，ID 为 d 的某汽车轨迹段 T_d 可以表示为有序时间序列：

$$T_d = \{P_1, P_2, \dots, P_i, \dots, P_n\} \quad (1)$$

$$P_i = (X_i, Y_i)$$

4.1.2 车辆道路行为

考虑相邻时间序列点的坐标变换，每一个时间序列点都对应一个车辆道路行为：直行、转弯、变道、静止。

$$DA(P_i) \in \{Str, Tur, Cha, Sta\} \quad (2)$$

具体定义如下：

1. 若连续 N 个时间序列点满足：

$$X_i = X_{i+1} = \dots = X_{i+N-1} \quad (3)$$

$$Y_i = Y_{i+1} = \dots = Y_{i+N-1}$$

则在时间段 T_N 内的时间序列点的车辆道路行为静止，即：

$$DA(P_i, P_{i+1}, \dots, P_{i+N-1}) = Sta \quad (4)$$

2. 定义位置坐标中相邻两个时间序列点的变化斜率：

$$K_i^{i+1} = \frac{Y_{i+1} - Y_i}{X_{i+1} - X_i} (X_{i+1} \neq X_i) \quad (5)$$

若连续 N 个时间序列点满足：

$$K_i^{i+1} = K_{i+1}^{i+2} = \dots K_{i+N-2}^{i+N-1} \quad (6)$$

则在时间段 T_N 内的时间序列点的车辆道路行为直行，即：

$$DA(P_i, P_{i+1}, \dots, P_{i+N-1}) = Str \quad (7)$$

3. 若在连续的 N_1 、 N_2 、 N_3 ($N_1 + N_2 + N_3 = N$) 个时间序列点中满足:

$$K_i^{i+1} = K_{i+1}^{i+2} = \dots K_{i+N_1-2}^{i+N_1-1} \neq K_{i+N_1-1}^{i+N_1} = K_{i+N_1}^{i+N_1+1} \dots = K_{i+N_2-2}^{i+N_2-1} \neq K_{i+N_2-1}^{i+N_2} = K_{i+N_2}^{i+N_2+1} = \dots = K_{i+N_3-1}^{i+N_3}$$

则在时间段 T_N 内的时间序列点的车辆道路行为变道, 即:

$$DA(P_i, P_{i+1}, \dots, P_{i+N-1}) = Tur \quad (8)$$

其中 N 为车辆道路行为最小判定采样点长度, T_N 为车辆道路行为最小判定周期, 其大小取决于具体数据集数据特征。

4.1.3 车辆直行运动状态

认为车辆在直行过程中的运动状态分为减速、匀速、加速三种。

$$Str \in \{Str_d, Str_c, Str_a\} \quad (9)$$

对于直行的有序时间序列:

$$\begin{aligned} T &= \{P_m, P_{m+1}, \dots, P_n\} \\ DA(P_m, P_{m+1}, \dots, P_n) &= Str \end{aligned} \quad (10)$$

其有序逐差距离序列为:

$$\begin{aligned} \Delta T &= \{\Delta P_1, \Delta P_2, \dots, \Delta P_t, \dots, \Delta P_{n-m}\} \\ \Delta P_t &= |P_{m+t-1} - P_{m+t}| \end{aligned} \quad (11)$$

每个采样时间内的平均速度为:

$$\bar{v}_t = \frac{\Delta P_t}{\Delta t} \quad (12)$$

对于直行的有序时间序列中的每一个序列点, 定义其与上一个时间序列点之间轨迹的平均速度和下一个时间序列点之间轨迹的平均速度的绝对变化误差:

$$|\Delta v_t| = \left| \frac{\bar{v}_{t+1} - \bar{v}_t}{\bar{v}_{t+1} + \bar{v}_t} \right| \quad (13)$$

若绝对变化误差小于速度变化误差阈值 γ , 则认为该时间序列点集合为匀速状态, 否则为加速或减速状态。

4.1.4 减速制动距离

在车辆直线行驶的过程中，司机往往会因为一些道路行为决策（如因红灯进行减速）或突发意外情况（如遭遇交通事故进行减速）对处在匀速运动状态中的车辆进行制动减速，车辆由匀速状态减速至静止状态的距离为减速制动距离，对于离散的减速有序时间序列：

$$T_{down} = \{P_m, P_{m+1}, \dots, P_n\} \quad (14)$$

在本题数据集中，其满足：

$$\begin{aligned} DA(P_{n+1}, P_{n+2}, \dots, P_{n+N_1}) &= Sta \\ DA(P_{m-N_2}, P_{m-N_2+1}, \dots, P_{m-1}) &= Str_c \\ DA(P_m, P_{m+1}, \dots, P_n) &= Str_d \end{aligned} \quad (15)$$

即车辆由均速运动状态开始减速，并逐渐减速至静止状态，经检验，本题数据中的减速事件发生场景均为车辆在路口遇到了红灯进行减速制动。对于一段离散的减速有序时间序列，其总制动距离为：

$$H_{down} = \sum_{i=m}^n |P_i P_{i+1}| \quad (16)$$

对于某路口固定范围内所有车辆发生的减速事件，定义该路口车辆平均减速制动距离：

$$\bar{H}_{down} = \frac{\sum H_{down}}{N_{down}} \quad (17)$$

其中 N_{down} 为该路口发生的总减速事件数。

4.1.5 交通路口简化模型

一个符合交通标准的交通路口包含许多要素，如红绿灯设计、虚实线设计、车道数量、路口类型，基于本题提供的数据，其可简化为两种交通路口模型：

(1) 双车道直行路口

仅考虑双车道，无需考虑路口为直行路口或丁字路口或十字路口，两车道均可直行，仅关注管理直行的红绿灯。车辆在双车道内在直行红绿灯的管理下经过路口，经过红绿灯前后一段距离内可以进行变道，在两个车道内均可因为红灯进行减速停车等待。示意图如图 1：

(2) 双车道左转路口

仅考虑双车道，无需考虑路口为丁字路口或十字路口，仅关注管理路口左转的红绿灯。车辆在进入左转车道后在左转红绿灯的管理下进行左转弯，在转弯前和转弯后的一段距离内可以进行变道，在左转车道内可以因为红灯进行减速停车等待。示意图如图 2：

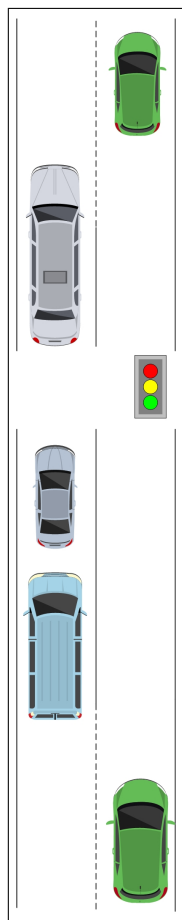


图 1: 直行模型示意图

4.1.6 红绿灯决策区域

车辆在驶入距离红绿灯较近的区域时，车辆驾驶者往往会根据路口情况的不同从而采取不同的车辆行驶决策，如红灯即将结束决策进行减速等待红灯、绿灯尚有剩余且路口无等待车辆决策进行加速通过绿灯等等。我们将车辆驾驶者需要由于红绿灯进行交通决策的交通区域定义为红绿灯决策区域。

(1) 历史等待峰值

历史等待峰值区域是指该路口数据集中因红灯而静止进行等待的车辆在车道中占据的最大区域，其宽度为参与红绿灯决策的车道总宽度，其长度为历史等待峰值 H_{max} 。

历史等待峰值区域在一定程度上反应了该路口的车流情况和红绿灯周期，若历史等待峰值过长，则说明该交通路口车流量较大，或红灯周期设置过长；若历史等待峰值过短，则说明该交通路口车流量较少，或者红绿灯总周期设计过短。

通过对历史等待峰值的数据分析，有助于了解特定时间段内交通路口的交通情况，判断交通标志设置的合理性，进而为交通规划优化提供数据支持。

(2) 简化红绿灯决策区域

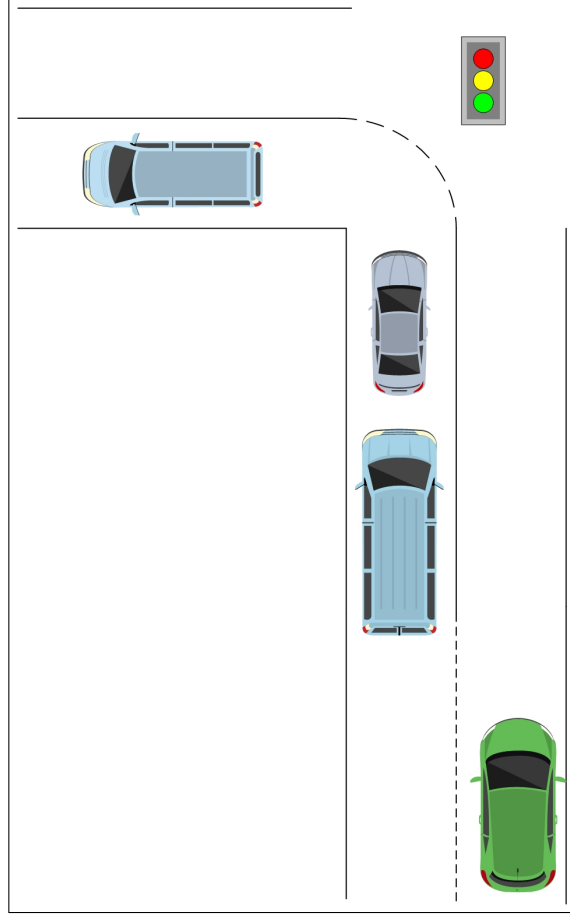


图 2: 左转模型示意图

考虑双车道模型，其红绿灯决策区域由历史等待峰值区域和减速制动区域组成。

历史等待峰值区域的长度 H_{max} 通过分析数据集的最值得到。当数据集的信息为全部车辆信息时，分析单位时间点内处于静止状态的车辆位置信息，即可获得路口红绿灯位置信息和历史等待峰值区域的位置坐标。

减速制动区域在历史等待峰值区域之后，其宽度同样为参与红绿灯决策的车道总宽，其长度为前文由数据集车辆减速时间序列计算得出的路口车辆平均减速制动距离 $\bar{H}_{down}$ 。

简化红绿灯决策区域由历史等待峰值区域和减速制动区域的总区域按照一定的误差比例系数扩大得到：

$$S_{rg} = (S_{max} + S_{down}) * (1 + \beta) \quad (18)$$

其决策区域示意图如图 3。

4.1.7 红绿灯离散时间序列

(1) 红绿灯简化模型

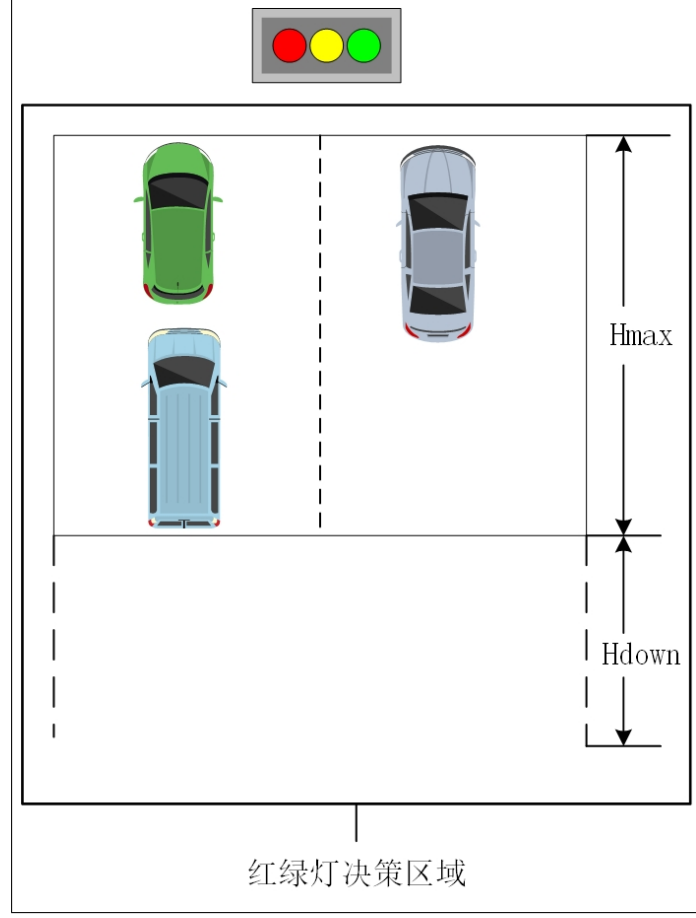


图 3: 红绿灯决策区域示意图

交通路口的红绿灯包含多个要素，如红绿灯的类型，是否包含黄灯，红绿灯总周期，周期是否发生变化，红绿灯时间在总周期中的占比等待。

本题考虑红绿灯简化模型，即红绿灯的类型为普通红绿灯，仅考虑管理左转和直行的红绿灯，不考虑黄灯。红绿灯的总周期 T_{sum} 由红灯时间 T_r 和绿灯时间 T_g 构成。即一个完成的红绿灯周期为：

$$T_{sum} = T_r + T_g \quad (19)$$

(2) 红绿灯状态离散时间序列

现考虑某路口的红绿灯对象，在时间采样间距 Δt , 时间段长度 T_{all} 的时间序列中，每一个时间序列点中的红绿灯对象都有唯一特征值，包括红灯、绿灯、由红灯转变为绿灯、由绿灯转变为红灯，即：

$$State(L_i) \in \{Red, Gre, RtG, GtR\} \quad (20)$$

在时间采样间距 Δt , 达到较小值时，红绿灯对象的特征值集合元素应当仅包含红灯或绿灯两种特征值，在本题研究对象中，将特征值精确缩小至红灯或绿灯即难实现，也无必要，红绿灯状态之间的有向转变特征也可作为红绿灯周期的确定提供支持。

(3) 红绿灯对象特征值判定

时间序列中每个红绿灯对象的特征值取决于红绿灯决策区域中车辆道路对象及其车辆道路行为，具体判定规则如下：

1. 当红绿灯决策区域中无车辆，此时无有效信息对特征值进行判定，即：

$$\begin{aligned} S_{rg} &= \emptyset \\ State(L_i) &= Null \end{aligned} \quad (21)$$

2. 当红绿灯决策区域中有车辆，且有车辆的车辆道路行为为静止，则此时红绿灯对象特征值为红灯，即：

$$\begin{aligned} \exists P_j \in S_{rg} \wedge DA(P_j) &= Sta \\ State(L_i) &= Red \end{aligned} \quad (22)$$

进一步的，观察静止车辆在接下来 N^* 个时间序列点中的车辆道路行为，若由静止变为直行加速，则下一个时间序列点的红绿灯对象特征值为红灯转变为绿灯，即：

$$\begin{aligned} \exists N^* \wedge AD(P_{j+1}, P_{j+2} \dots P_{j+N}) &= Str_a \\ State(L_{i+1}) &= RtG \end{aligned} \quad (23)$$

其中 N^* 为最小车辆道路行为判定数。

3. 当红绿灯决策区域中有车辆, 且全部车辆的车辆道路行为均为直行减速速，则此时红绿灯对象特征值为红灯, 即：

$$\begin{aligned} \forall P_j \in S_{rg} \wedge AD(P_j) &= Str_d \\ State(L_i) &= Red \end{aligned} \quad (24)$$

4. 当红绿灯决策区域中有车辆, 且全部车辆的车辆道路行为均为直行匀速，则此时红绿灯对象特征值为绿灯, 即：

$$\begin{aligned} \forall P_j \in S_{rg} \wedge AD(P_j) &= Str_c \\ State(L_i) &= Gre \end{aligned} \quad (25)$$

由于决策区域的存在，认为车辆驾驶者在面对相同道路情况时将会作出相同的道路决策，所以不存在在红绿灯决策区域中既有车辆直行减速，又有车辆直行匀速的情况。此外，经数据检验，本题中的直行加速决策全部出现在车辆静止之后，即不存在绿灯即将结束，车辆驾驶者进行加速通过绿灯的情况。

此外，由于判定中存在对下一时间序列点的红绿灯对象特征值判定，以上判定规则仅适用于该时间序列点的特征值未进行判定或判定值为 $Null$ 的前提下。

4.1.8 确定理想时间序列点算法

红绿灯由绿灯转变为红灯的时间序列点是求取红绿灯周期的关键，其可能发生在红绿灯决策区域中无车的情况下，也可能在绿灯即将变为红灯的前后时间序列中，车辆驾驶者均进行直行减速的道路决策。针对此问题，我们在上述获得的红绿灯状态特征值离散时间序列的基础上，设计算法求取理想的绿灯转变为红灯时间序列点。

(1) 遗传算法

遗传算法（Genetic Algorithm, GA）是一种模拟自然选择和遗传机制的优化算法，常用于解决搜索和优化问题。算法基于达尔文的进化论思想，通过模拟自然界中的遗传、变异、选择和适应度等过程，不断进化出符合问题要求的解。

遗传算法通常包括以下基本步骤：

1. 初始化种群：随机生成初始种群，种群中的每个个体代表问题的一个可能解。
2. 评估适应度：计算种群中每个个体的适应度，即解决问题的能力或优劣程度。
3. 选择操作：根据个体的适应度，选择适应度较高的个体作为父代，用于繁殖下一代。
4. 交叉操作：通过交叉操作，从父代个体中产生新的子代个体。交叉操作类似于生物的交配过程，可以产生新的个体组合。
5. 变异操作：对新生成的子代个体进行变异操作，引入一定程度的随机性，以增加种群的多样性。
6. 替换操作：根据一定的策略，用新生成的子代个体替换原来的父代个体，形成新一代种群。
7. 重复进化：重复进行选择、交叉、变异和替换操作，直到达到停止条件（如达到最大迭代次数或找到满足要求的解）为止。遗传算法具有较强的全局搜索能力和较好的鲁棒性，适用于复杂的优化问题和搜索空间较大的情况，算法的性能和效率受到参数设置、编码方式、交叉和变异算子设计等因素的影响。

(2) 基于遗传算法确定理想时间序列点

考虑特征值离散时间序列，建立基于遗传算法的理想时间序列点算法，具体如下：

优化目标：寻找序列中最适合作为红灯转变为绿灯特征值的时间序列点

编码方式：将离散时间序列作为字符串进行编码

变异方式：随机选取对应特征值为 *Null* 或 *Gre* 的字符并将其特征值更改为 *RtG*

交叉方式：对于两个变异不同的字符串对象，对于其中一方字符对应特征值为 *RtG* 而另一方不是的情况，交叉结果为对应特征值为 *RtG* 的字符

适应度：对与包含特征对象为 *Null* 或 *RtG* 的字符串对象，其被红绿灯转换节点分割为有限个红绿灯周期，计算其周期长度的均方误差作为适应度，均方误差越小，说明分割得到的周期越小，适应度也就越强

其具体流程图如图 4。

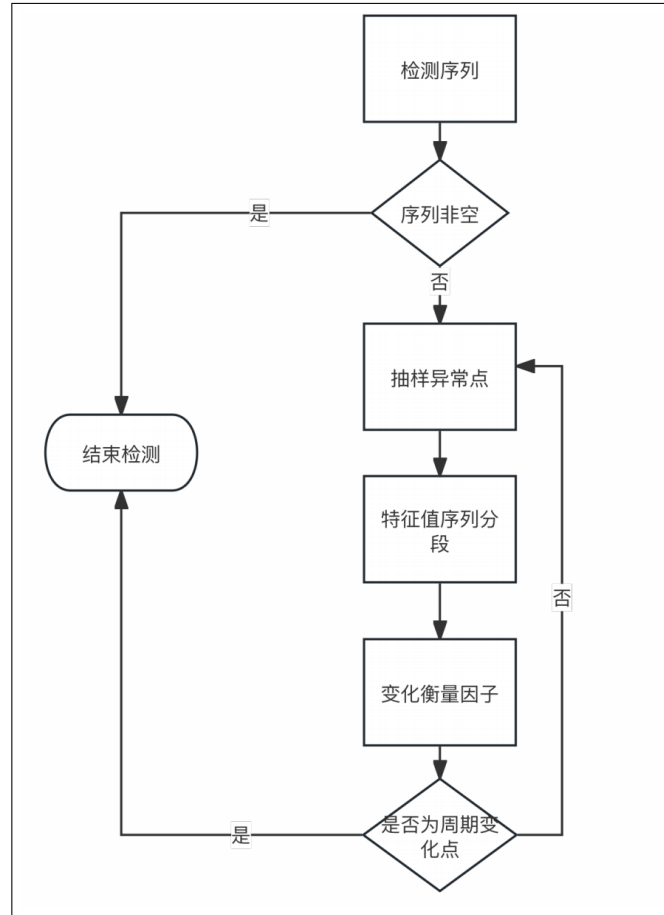


图 4: 流程图

4.1.9 问题一模型的求解

五个路口的制动距离如下：

路口	A1	A2	A3	A4	A5
制动距离	2.96	3.47	3.33	8.24	7.95

附件 5 数据集计算出的红绿灯决策区域的坐标如下：

决策区域	(x_1, y_1)	(x_2, y_2)	(x_3, y_3)	(x_4, y_4)
坐标	(1.53, -37.77)	(4.57, -7.37)	(0.43, -6.96)	(-2.61, -37.36)

最终求得红绿灯周期为：

路口	A1	A2	A3	A4	A5
红灯时长 (秒)	70	54	82	70	64
绿灯时长 (秒)	34	31	35	28	31

4.2 问题二模型的建立与求解

考虑在实际情况中，只有部分用户使用该公司的产品，即只能获取部分样本车辆的行车轨迹。同时，受各种因素的影响，轨迹数据存在定位误差。样本车辆的比例、车流量、定位误差等因素都会对上述模型的最终计算结果产生影响，因此我们建立误差灵敏度分析模型，来衡量和判断上述误差因素对模型参数的影响，并根据误差信息进行数据优化，对最终结果进行修正。

4.2.1 误差灵敏度分析模型

(1) 车流量

车流量通常指的是某段时间内通过某个地点或道路的车辆数量。它是衡量交通繁忙程度和道路使用情况的重要指标之一。车流量数据可以用于交通规划、道路设计、交通管理等方面。本题中对于车辆的位置信息采集均在路口前后固定车道范围内进行，因此将当前时间序列点采集范围内的全部车辆数量用于衡量车流量，即：

$$P_{all} = \{P_1, P_2, \dots, P_i, P_{i+1}, \dots, P_j, P_{j+1}, \dots, P_n\} \in S_{all} \quad (26)$$

$$TV = |P_{all}|$$

其中 S_{all} 为车辆位置信息采集区域， P_{all} 为当前时间序列点内全部车辆位置信息集合， TV 为车流量。

(2) 样本占比

考虑不定采样操作集 SA ，在给定的样本数据集中，其时间序列实际是真实完整时间序列的不定采样子集，即：

$$P_{sam} = SA(P_{all}) \quad (27)$$

操作集的大小即为样本数据集的大小，即可计算得样本车辆的比例：

$$\lambda = \frac{|P_{sam}|}{|P_{all}|} = \frac{|SA|}{TV} \quad (28)$$

(3) 定位误差

在样本数量集中，车辆的位置信息由当前位置 x 坐标、当前位置 y 坐标决定。由于在测量的过程中因为随机扰动等各种因素，可能会出现精度误差、时间误差等各种误差。位置信息的误差在一定程度上影响了车辆道路行驶状态的判定、红绿灯决策区域的判定等等。

在本题中，仅考虑定位误差带来的影响，每个车辆位置时间序列点的坐标存在微小误差波动，且认为该波动分布符合标准高斯分布，即：

$$P_i\{X_i + \epsilon_1, Y_i + \epsilon_2\} \quad (29)$$

$$\epsilon \sim \mathcal{N}(0, 1)$$

该定位误差影响车辆道路行为的判定和车辆是否处于红绿灯决策区域的判定，前者即在全部数据集上以符合正态分布的微小误差体现，后者由于引入误差比例系数而已经被消除。

(4) 有效信息损失量

有效信息是指能够为红绿灯特征值的判定提供支持的车辆状态时间序列点，有效信息越多，对于当前红绿灯特征值的判断也就越准确。

当操作集的大小逐渐增大时，样本数据集越接近真实数据集，损失的有效信息也就越少；当车流量逐渐增大时，真实数据集的时间序列长度也就越长，损失的有效信息也就越多。也就是说，车流量和样本车辆比例以及定位误差影响着有效信息的损失量，即：

$$Loss(\lambda, TV) = K \frac{TV}{\lambda} + \epsilon^* \quad (30)$$

$$\epsilon^* \sim \mathcal{N}(0, 1)$$

其中 $Loss(\lambda, TV)$ 是由样本车辆比例、车流量共同决定的有效信息损失量衡量函数， K 为比例系数。

4.2.2 基于误差分析的数据优化

对于损失一定有效信息的样本数据集，其主要影响体现在红绿灯特征值的判定上。即对于任意的离散时间红绿灯时间序列点，其车辆道路行为数据集分为样本集和损失集：

$$T_{ture} = \{\{P_1, P_2, \dots, P_n\}, \{P_1^*, P_2^*, \dots, P_m^*\}\} \quad (31)$$

对于每一个损失集通过 $Loss(\lambda, TV)$ 函数可以计算损失量，并根据损失量的大小进行判定，若在一定小的范围内，认为接受损失，否则，按照加权系数随机引入车辆道路行为信息，对样本数据集进行补充。在完成数据优化后，再进行该点的红绿灯特征值判定，即：

$$DE(\theta^*, Loss) = \{k_1, k_2, k_3\} \quad (32)$$

$$T_{add} = \{k_1 * \{Sta\}, k_2 * \{Str\}, k_3 * \{Cha, Tur\}\}$$

其中 $DE(\theta^*, Loss)$ 为损失判定函数， θ^* 为损失判定阈值，其根据 $Loss$ 的大小返回补充数据集的加权系数， k_1 为补充的车辆道路行为为静止的数据集系数， k_2 为补充的车辆道路行为为直行的数据集系数， k_3 为补充的车辆道路行为为转弯或变道的数据集系数， T_{add} 为补充的车辆道路行为离散时间序列。

4.2.3 问题二模型的求解

最终求得红绿灯周期为：

路口	B1	B2	B3	B4	B5
红灯时长 (秒)	78	81	72	79	92
绿灯时长 (秒)	26	30	38	39	22

4.3 问题三模型的建立与求解

信号灯周期有可能发生变化，在周期发生变化后，在该路口车流量稳定的情况下，因为红绿灯而发生车辆道路行为决策的数据整体会产生一定波动，通过对于这种波动的检测，可以对红绿灯周期变化发生的时间点进行确定。

4.3.1 变化检测密度抽样

密度抽样是一种基于数据点密度的异常检测方法，通常用于连续数据集。其基本思想是在数据集中选择具有低密度的数据点作为候选异常点。密度抽样的步骤如下：

1. 密度估计：首先，需要对数据集中每个数据点的密度进行估计。常用的密度估计方法包括基于距离的方法（如K近邻密度估计）和基于核函数的方法（如高斯核密度估计）。

2. 选择阈值：根据密度估计结果，选择一个适当的阈值。密度低于阈值的数据点被认为是异常点的候选。

3. 密度抽样：从密度低于阈值的数据点中进行抽样，作为最终的异常点。本数据集的抽样对象为红绿灯特征值时间序列，并根据抽样出的异常点将时间序列进行分段。即：

$$L = \{L_1, L_2, \dots, L_{out}, \dots, L_m, L_{m+1} \dots L_n\} \quad (33)$$

其中 L_{out} 为异常点。

4.3.2 分段数据的变化检测

在前文中，我们定义并计算了简化红绿灯决策区域，对于分段后的红绿灯特征值时间序列，我们重新计算子数据集的决策区域位置，并与相邻时间序列数据集进行比较，若在一定范围内保持不变，则认为该分段数据的红绿灯周期并未发生变化，即：

$$C = \left| \frac{\frac{S_2 - S_1}{S_1} - \frac{S_3 - S_2}{S_2}}{\frac{S_2 - S_1}{S_1} + \frac{S_3 - S_2}{S_2}} \right| \quad (34)$$

其中 C 为变化衡量因子， S_i 为决策区域面积。

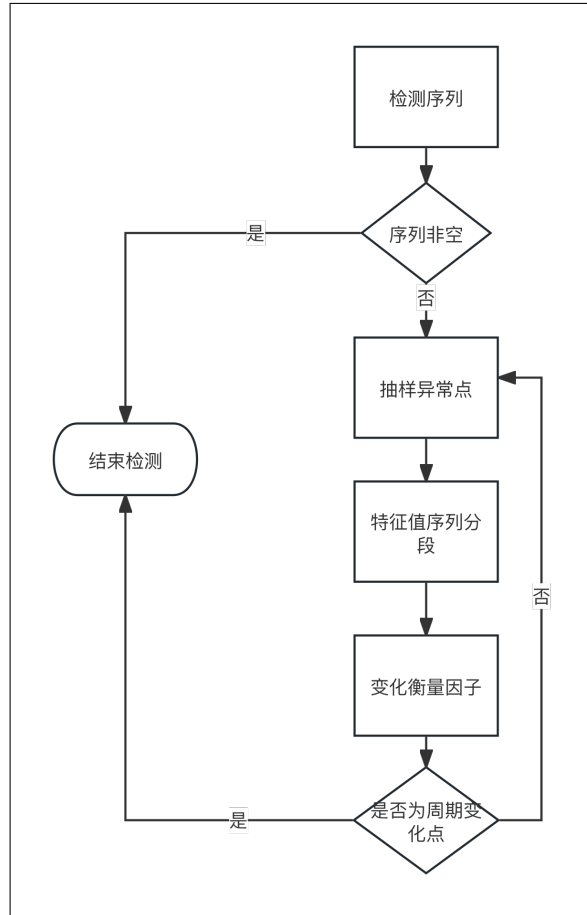


图 5: 检测示意图

4.3.3 遍历确定周期变化点

对于抽样得到的异常点，若分段后得到的分段时间序列的红绿灯决策区域未发生较大变化，则将该异常点从检测序列中排除，并在剩余的序列集中重新进行密度抽样，直到发现周期变化点活检测序列集为空，其具体流程图如图 5

4.4 问题三模型的求解

最终求得红绿灯周期为：

路口	C1	C2	C3	C4	C5	C6
周期 1 红灯时长 (秒)	57	67	81	77	56	77
周期 2 绿灯时长 (秒)	32	26	27	30	31	25
周期切换时刻	1214	1845	2011	986	775	1423
周期 2 红灯时长 (秒)	46	46	65	60	35	62
周期 2 绿灯时长 (秒)	42	40	32	36	45	37

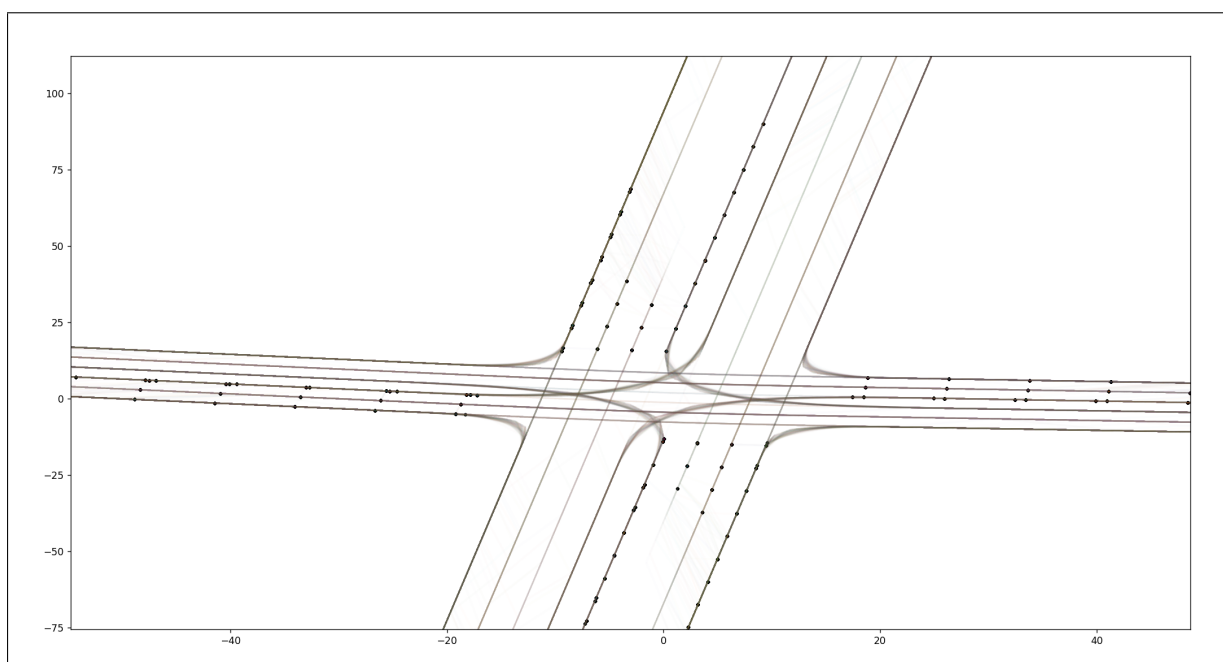


图 6: 可视化路口四方向示意图

4.5 问题四模型的建立与求解

前三问中的车辆位置信息数据集仅构成交通路口的一个方向，通过对问题四中的数据集进行初步分析和可视化，其提供的车辆位置信息构成交通路口四方向，其一段时间内的可视化图像如图 6。

4.5.1 四方向交通路口模型

考虑四方向交通路口模型，其每个方向的道路含八车道，四个方向的车均只进行左转，并由左转红绿灯进行管理，四个方向由两个红绿灯进行管理，具体示意图如图 7。

此时，当其中一红绿灯为红灯时，允许其中两车道进行左转；另一红绿灯对应为绿灯，静止另外两车道左转。

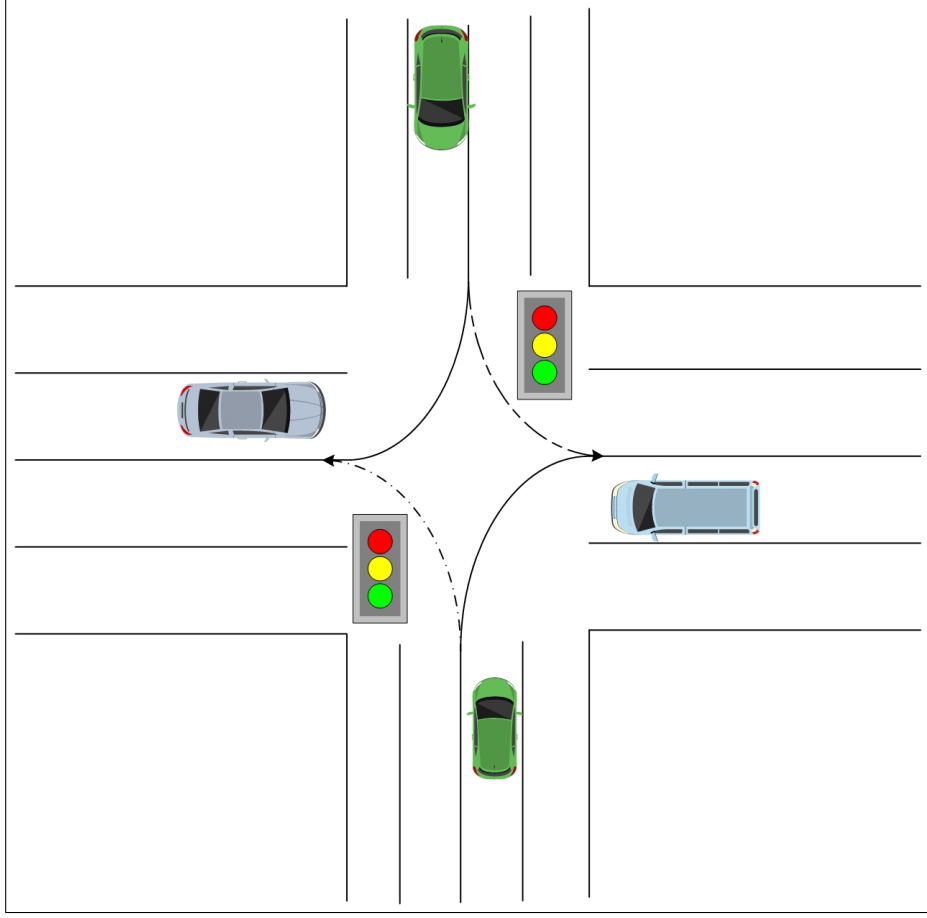


图 7: 四路口示意图

4.5.2 四方向车辆位置时间序列

考虑四方向车辆位置构成的时间序列，将其考虑为两个红绿灯对象和两个车辆道路行为时间序列对象，并用于确定最终的红绿灯离散时间特征值序列，即：

$$L_1, L_2 \in \{Red, Gre\} \quad (35)$$

$$L_1 = \{P_1, P_2, \dots, P_n\}$$

$$L_2 = \{P_1, P_2, \dots, P_m\}$$

将其转化为红绿灯离散时间特征值序列，并根据问题一的建模进行求解。

4.5.3 问题四模型的求解

最终求得红绿灯周期为：

路口	D
红灯时长 (秒)	119
绿灯时长 (秒)	72

参考文献

- [1] 赵肄江等.2024. 一种基于轨迹数据的红绿灯位置检测方法
- [2] Holland, J. H. 1975. Adaptation in Natural and Artificial Systems. University of Michigan Press.
- [3] Zhang, Y., Wang, J., & Liu, Y. 2018. Research on the Green Light Cycle of Traffic Signal Based on the Characteristics of Vehicle Trajectory Data.
- [4] Li, J., Zhang, H., & Wang, Y. 2019. Study on Traffic Signal Timing Based on Vehicle Trajectory Data.
- [5] Chen, S., Liu, B., & Zhang, H. 2017. Study on the Signal Control Method of Urban Intersection Based on Vehicle Trajectory Data.