基于遗传算法的交通信号灯控制管理

摘要

# 问题重述

# 问题分析

# 模型假设

1. 假设经中路和纬中路的所有路段具有一致的道路条件，即道路宽度、路面状况等，不因位置不同而产生显著的变化。
2. 假设每个交叉路口的每个相位的单位时间内通过的最大车辆数恒定，不随时间或交通状况发生变化。
3. 假设行人过街对车辆流量的影响忽略不计。
4. 假设交通模式在分析时段内保持稳定。
5. 假设所有车辆严格遵守交通规则，不会出现异常、违反交通规则的行为。

# 符号说明

# 模型建立与求解

## 5.1 问题一的模型建立与求解

### 5.1.1 数据预处理

本题数据量较大，在模型建立之前需要对数据进行数据预处理，发现数据中不存在重复值和缺失值，但是数据中存在大量车牌号为“无车牌”的数据，这些数据作为异常值处理。经统计，该异常值数据存在约为20万条，对于整个数据超800万占比较小，约占比0.025，因此可以将其剔除，对模型建立影响较小。

将经中路-纬中路的数据整理筛选出来。检查数据所包含的日期为2024年4月1日至2024年5月6日，考虑到节假日以及工作日对车流量的影响，将数据划分为工作日和节假日两类，便于后续的模型建立。查询日历，节假日中除了周末以外，还包括4月4日（清明节假期）以及5月1日至5月6日（劳动节假期）。根据调休规则，将4月7日（周日）和4月28日（周日）划分为工作日。具体区分见支撑材料一（工作日为True即为工作日，为False则为节假日）。

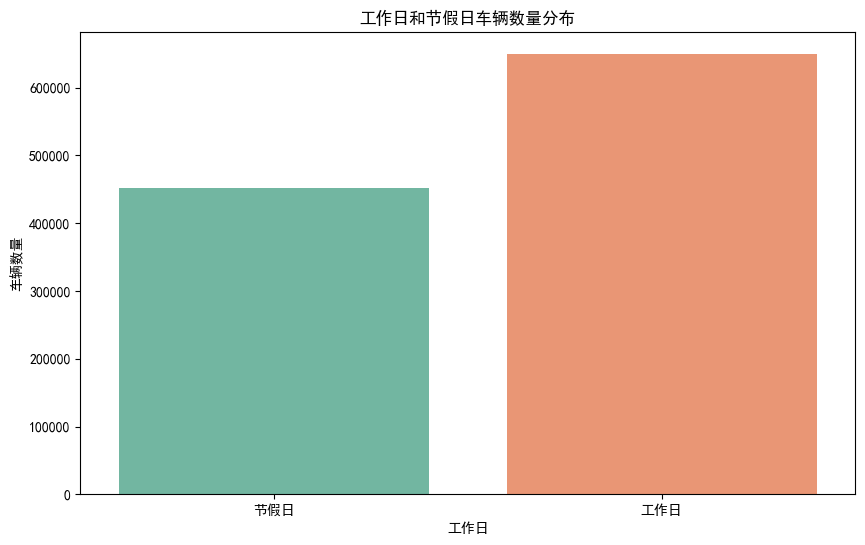
因此可以统计出节假日和工作日的车辆数量，如图1. 1所示：

图1. 1节假日总车流量和工作日总车流量对比图

随后统计每天的车流量数据，如图1. 2所示：

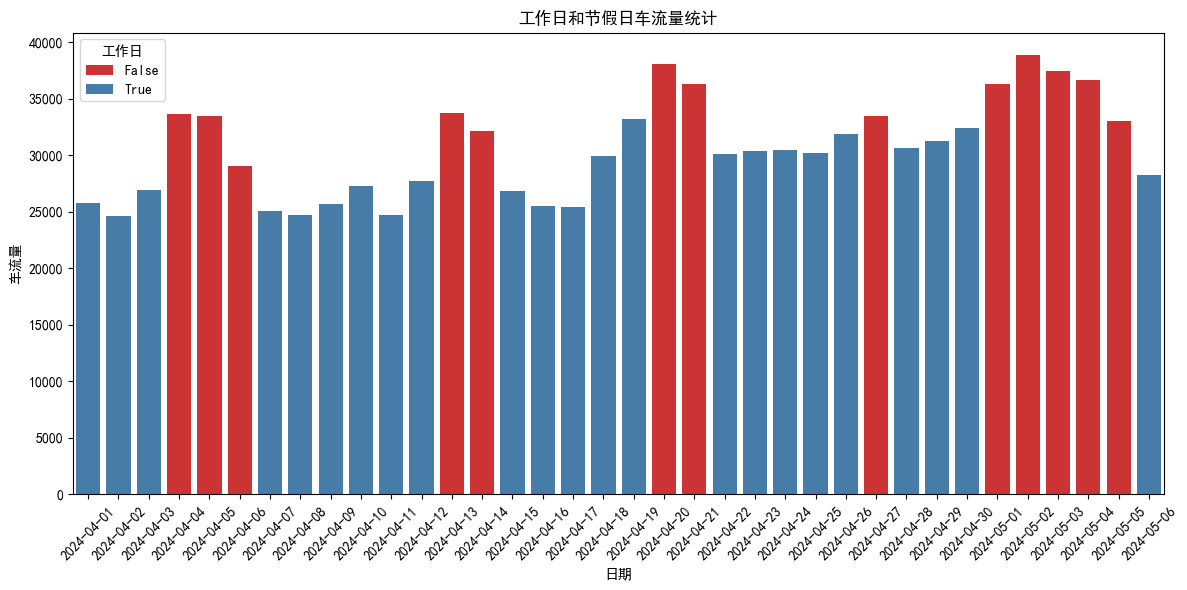


图1. 2经中路-纬中路每天的车流量数据

随后统计经中路-纬中路每个小时的车流量数据，如图1. 3和图1. 4所示：

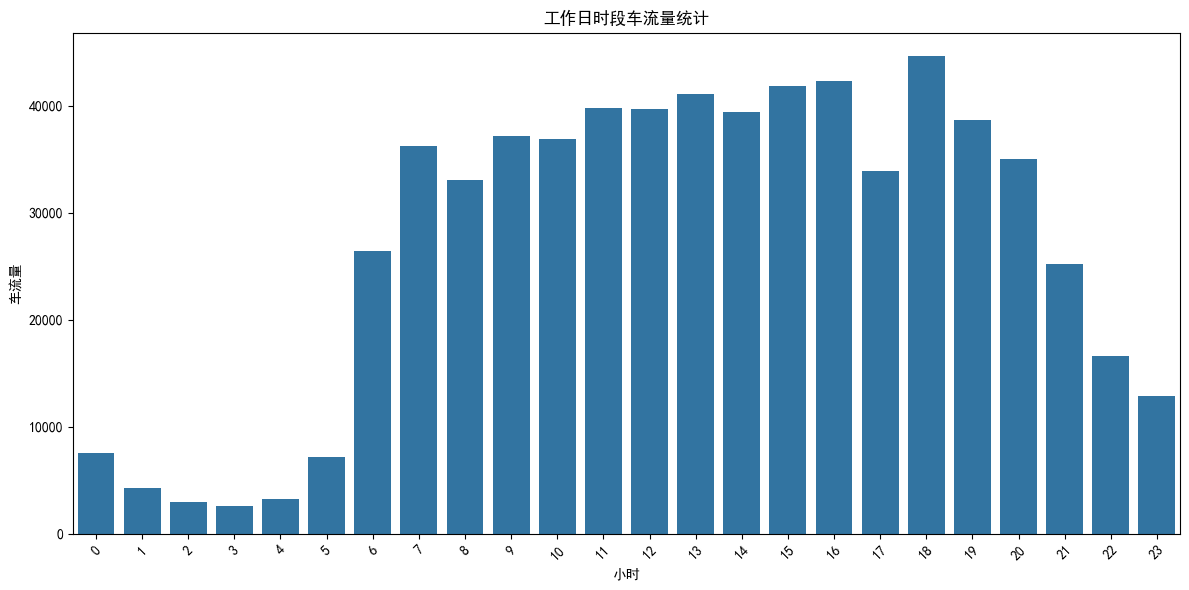


图1. 3经中路-纬中路工作日各小时车流量数据统计图

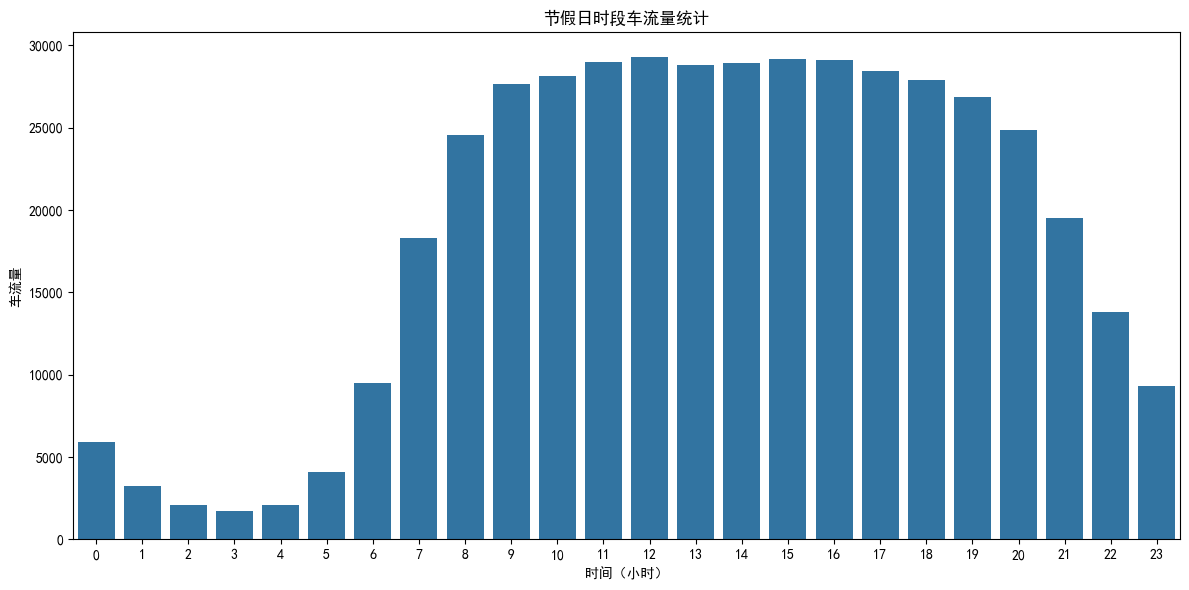


图1. 4经中路-纬中路节假日各小时的车流量数据统计图

由以上图片可以得到，工作日总车流量高于节假日总车流量，但在每一天的车流量上，节假日车流量普遍高于工作日车流量。针对一天中各个小时车流量数据，可以发现，无论是节假日还是工作日车流量普遍在0时至5时处于一天中车流量较低的时段，6时至21时普遍处于一天中车流量较高的时段，21时-23时车流量数据普遍处于中等时段。

因此我们可以得出，大体上对于经中路-纬中路时段可以划分为三个时段。

接下来统计出节假日和工作日各方向车流量折线统计图，如图1. 5和图1. 6所示：

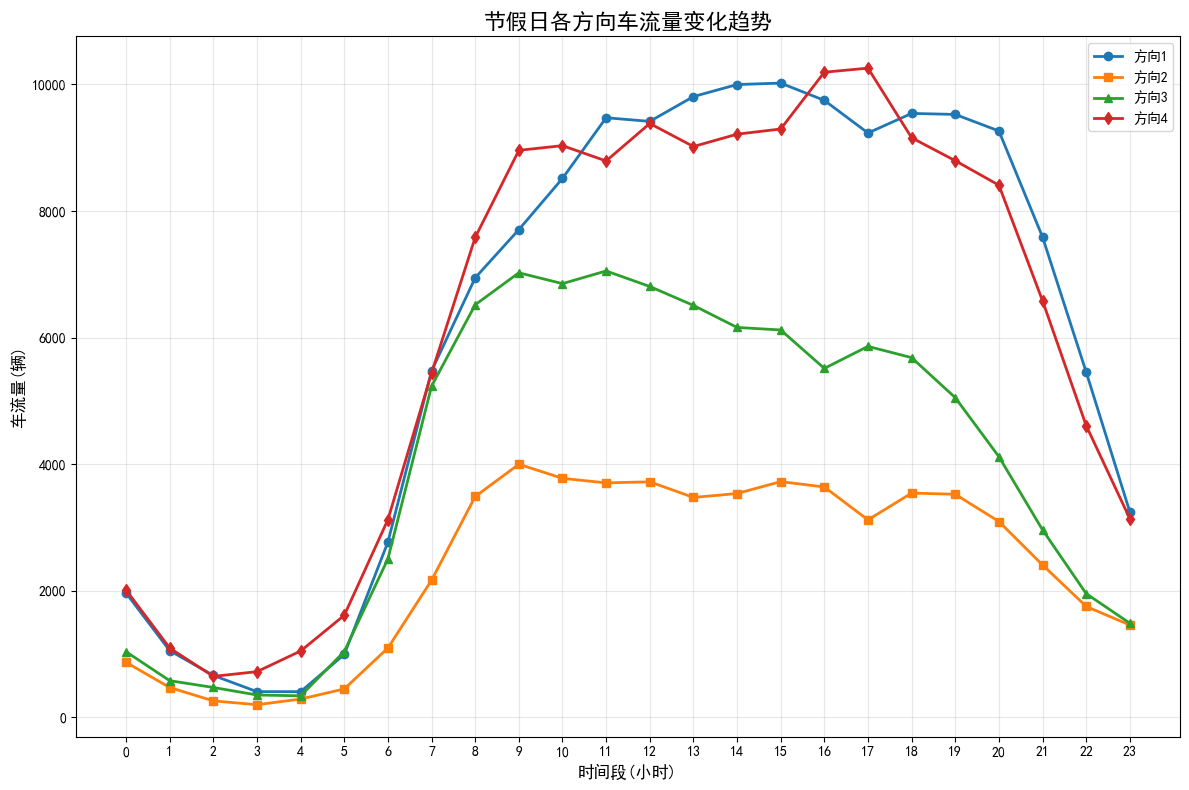


图1. 5节假日各方向车流量数据统计图

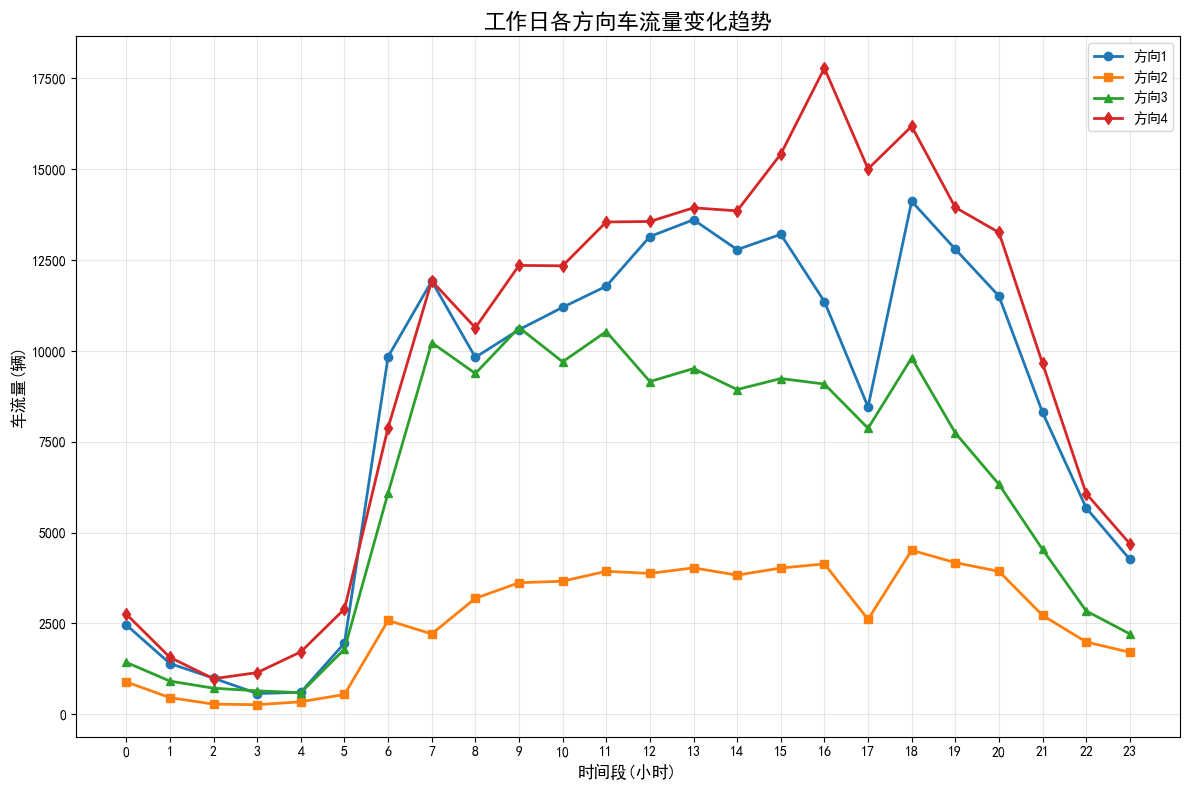


图1. 6工作日各方向车流量数据统计图

由图可知，节假日在12点以前整体上各方向车流量排名为：方向4＞方向1＞方向3＞方向2，在12点以后，整体上各方向车流量排名为：方向1＞方向4＞方向3＞方向2；工作日全天整体上车流量排名为：方向4＞方向1＞方向3＞方向2。由附件一可知方向1为由东向西，方向2为由西向东，方向3为由南向北，方向4为由北向南。因此，整体来看，节假日12点之前和工作日全天中，方向4的车辆较多；节假日12点以后方向1车辆较多。

### 5.1.2 K-means聚类算法模型的建立

K-means聚类是一种基于距离的非监督学习算法，目的是将数据划分为个簇，使得同一个簇内的数据点的距离最小化，而不同簇之间的差异最大化。根据本题目，我们建立以下目标函数：

 （1）

其中：为簇的数量，为第个簇，为数据点，为第个簇的中心，为数据点与簇的中心的欧式距离。

### 5.1.3 模型的求解

根据题目要求，我们需要将一天划分为若干个时段。上文数据预处理过程当中，我们大致可以将数据分为三个时段，通过K-means聚类算法求解就是取，但这并不一定是最优的值，我们需要根据肘部法则和轮廓系数结合来确定的取值。我们以肘部法则为主，通过轮廓系数辅助确定最终的值。

得到工作日和节假日的肘部法则图以及轮廓系数图，如图1. 7至图1. 10所示：

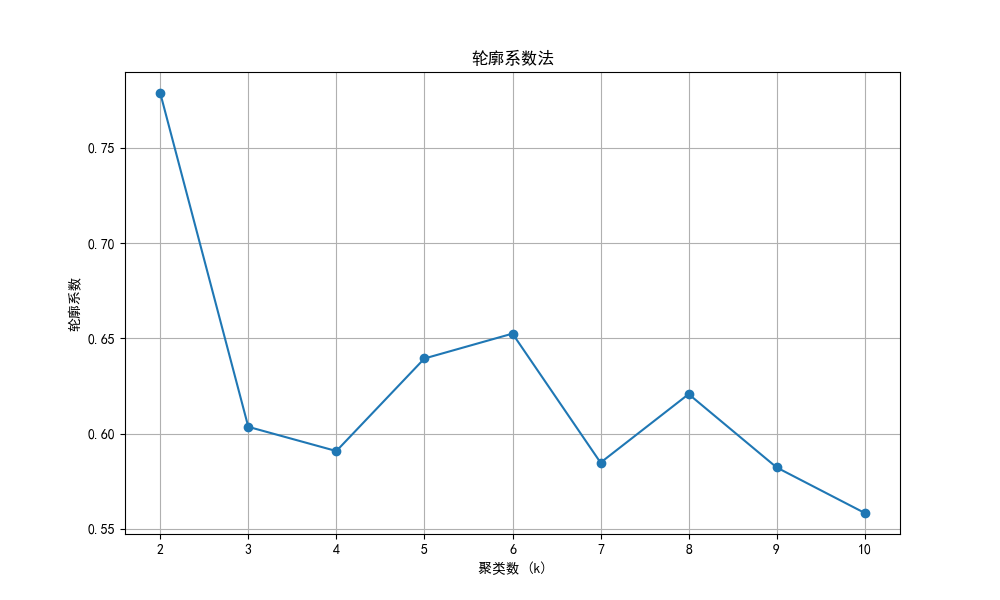


图1. 7工作日轮廓系数图像

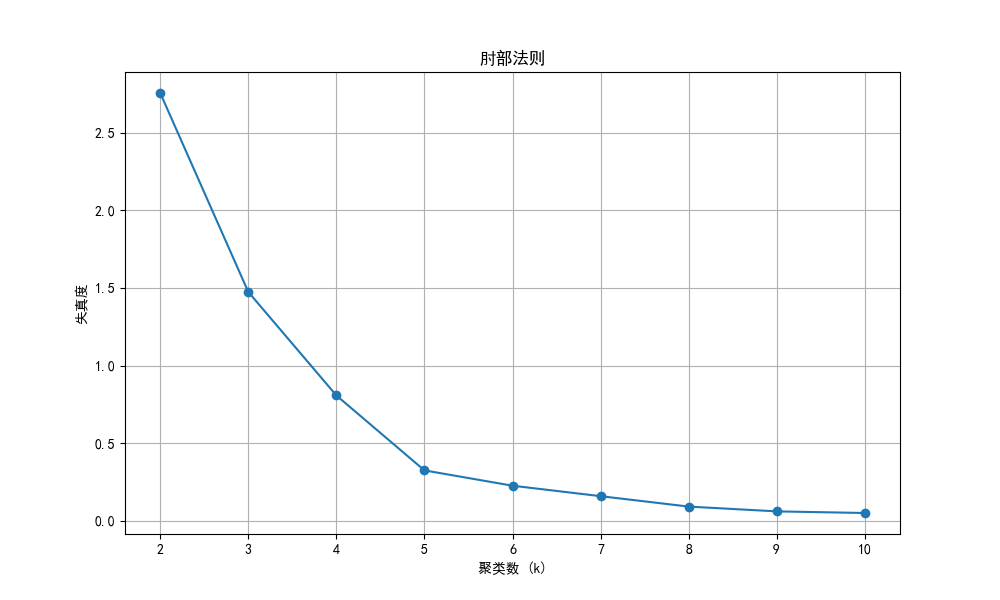


图1. 8工作日肘部法则图像

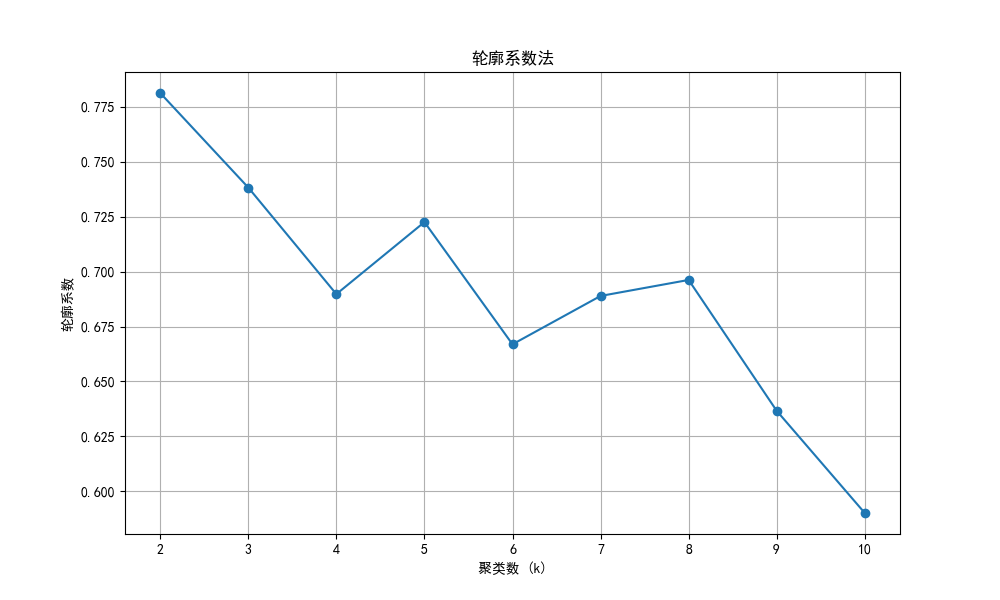


图1. 9节假日轮廓系数图像

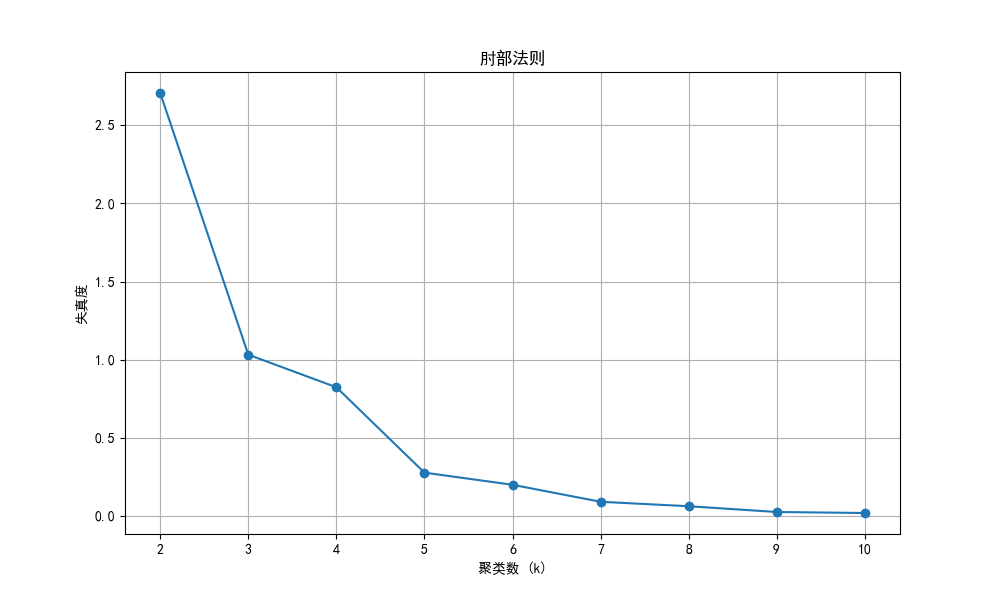


图1. 10节假日肘部法则图像

根据肘部法则和轮廓系数，我们可以更准确的确定工作日和节假日的最优值均为。因此，我们设定通过K-means聚类算法划分数据集，最终得到聚类结果，如图1. 11和图1. 12所示：

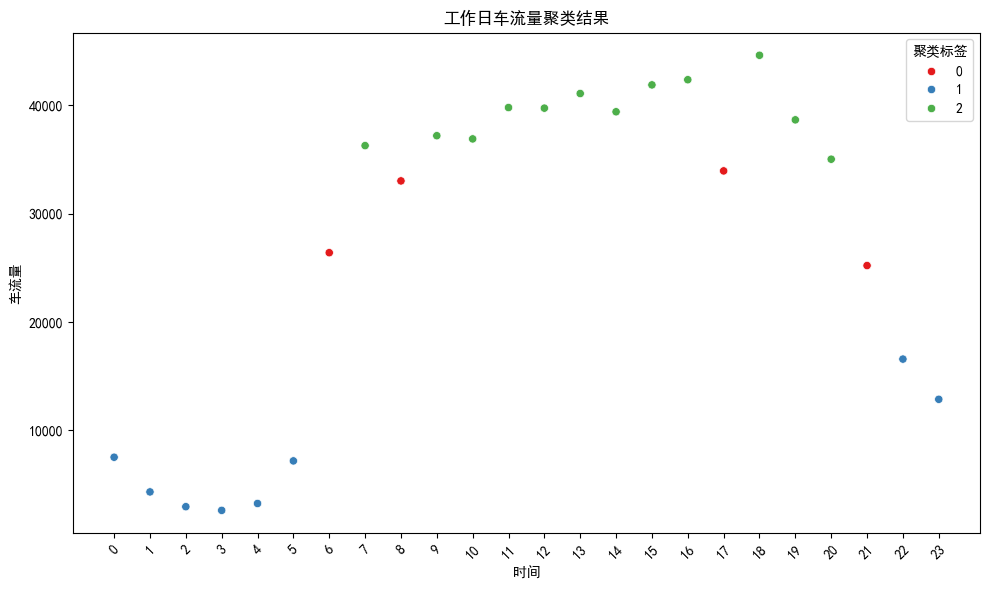


图1. 11工作日K-means聚类结果图

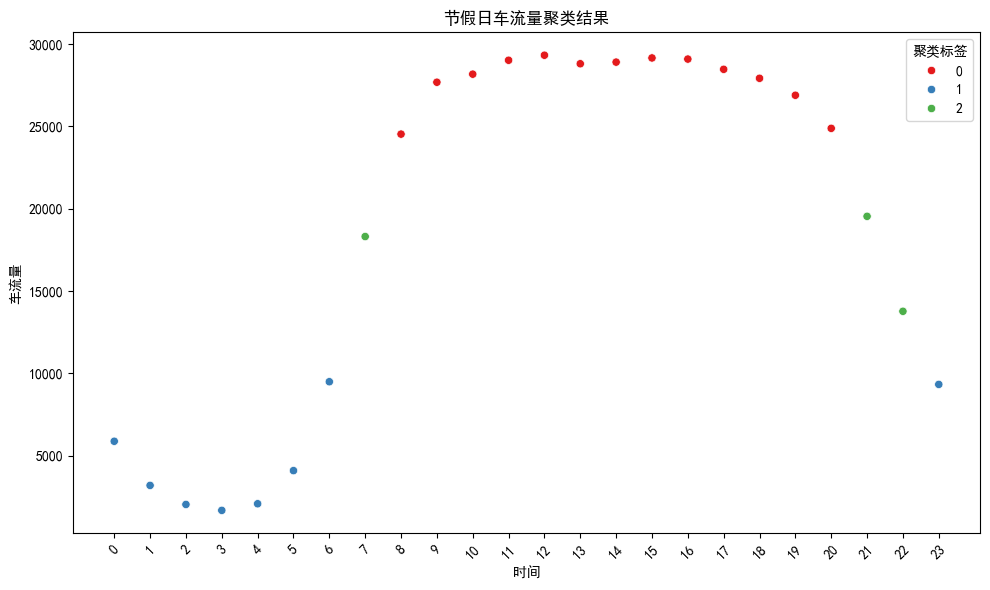


图1. 12节假日K-means聚类结果图

根据图片，我们根据聚类结果可以将工作日聚类标签进行排序为：标签2＞标签0＞标签1，因此，可以将标签2划分时段为高峰期，将标签0划分时段为平峰期，将标签1划分时段为低峰期；同理，节假日标签排序为：标签0＞标签2＞标签1，因此，可以将标签0划分为高峰期，标签2划分为平峰期，标签1划分为低峰期。

最终得到划分结果如下：

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 小时 | 方向1 | 方向2 | 方向3 | 方向4 | 聚类标签 | 时段 |
| 0 | 2456 | 888 | 1426 | 2753 | 1 | 低峰 |
| 1 | 1394 | 455 | 910 | 1563 | 1 | 低峰 |
| 2 | 991 | 276 | 715 | 976 | 1 | 低峰 |
| 3 | 568 | 261 | 643 | 1145 | 1 | 低峰 |
| 4 | 604 | 341 | 591 | 1717 | 1 | 低峰 |
| 5 | 1955 | 543 | 1791 | 2895 | 1 | 低峰 |
| 6 | 9843 | 2579 | 6093 | 7888 | 0 | 平峰 |
| 7 | 11921 | 2210 | 10228 | 11928 | 2 | 高峰 |
| 8 | 9824 | 3188 | 9379 | 10638 | 0 | 平峰 |
| 9 | 10584 | 3619 | 10644 | 12354 | 2 | 高峰 |
| 10 | 11201 | 3660 | 9701 | 12344 | 2 | 高峰 |
| 11 | 11782 | 3934 | 10534 | 13550 | 2 | 高峰 |
| 12 | 13149 | 3875 | 9157 | 13564 | 2 | 高峰 |
| 13 | 13612 | 4028 | 9515 | 13940 | 2 | 高峰 |
| 14 | 12789 | 3828 | 8937 | 13856 | 2 | 高峰 |
| 15 | 13210 | 4023 | 9240 | 15419 | 2 | 高峰 |
| 16 | 11360 | 4136 | 9088 | 17785 | 2 | 高峰 |
| 17 | 8459 | 2611 | 7870 | 15012 | 0 | 平峰 |
| 18 | 14120 | 4513 | 9808 | 16184 | 2 | 高峰 |
| 19 | 12801 | 4171 | 7743 | 13950 | 2 | 高峰 |
| 20 | 11509 | 3933 | 6328 | 13262 | 2 | 高峰 |
| 21 | 8305 | 2720 | 4535 | 9655 | 0 | 平峰 |
| 22 | 5679 | 1988 | 2842 | 6073 | 1 | 低峰 |
| 23 | 4262 | 1704 | 2210 | 4693 | 1 | 低峰 |

表1. 1工作日时段划分结果表

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 小时 | 方向1 | 方向2 | 方向3 | 方向4 | 聚类标签 | 时段 |
| 0 | 1964 | 871 | 1038 | 2012 | 1 | 低峰 |
| 1 | 1055 | 473 | 581 | 1096 | 1 | 低峰 |
| 2 | 664 | 262 | 474 | 649 | 1 | 低峰 |
| 3 | 408 | 202 | 355 | 724 | 1 | 低峰 |
| 4 | 408 | 291 | 342 | 1050 | 1 | 低峰 |
| 5 | 996 | 451 | 1042 | 1614 | 1 | 低峰 |
| 6 | 2771 | 1098 | 2506 | 3125 | 1 | 低峰 |
| 7 | 5473 | 2168 | 5233 | 5441 | 2 | 平峰 |
| 8 | 6944 | 3490 | 6518 | 7582 | 0 | 高峰 |
| 9 | 7705 | 3999 | 7023 | 8957 | 0 | 高峰 |
| 10 | 8514 | 3777 | 6852 | 9031 | 0 | 高峰 |
| 11 | 9472 | 3705 | 7052 | 8789 | 0 | 高峰 |
| 12 | 9414 | 3722 | 6807 | 9381 | 0 | 高峰 |
| 13 | 9806 | 3475 | 6508 | 9017 | 0 | 高峰 |
| 14 | 9996 | 3537 | 6160 | 9212 | 0 | 高峰 |
| 15 | 10019 | 3725 | 6120 | 9294 | 0 | 高峰 |
| 16 | 9746 | 3640 | 5513 | 10191 | 0 | 高峰 |
| 17 | 9231 | 3121 | 5861 | 10255 | 0 | 高峰 |
| 18 | 9540 | 3544 | 5683 | 9156 | 0 | 高峰 |
| 19 | 9524 | 3525 | 5051 | 8793 | 0 | 高峰 |
| 20 | 9263 | 3095 | 4119 | 8407 | 0 | 高峰 |
| 21 | 7588 | 2409 | 2963 | 6582 | 2 | 平峰 |
| 22 | 5456 | 1754 | 1956 | 4610 | 2 | 平峰 |
| 23 | 3248 | 1460 | 1489 | 3140 | 1 | 低峰 |

表1. 2节假日时段划分结果表

针对问题一，我们需要估计不同时段的各相位（包括四个方向直行、转弯）车流量，由于监控设备安装在停车线后方，因此并不知道车辆通过停车线后是左转、直行还是右转。因此我们需要通过粒子群优化算法对左转、直行和右转车辆的比例进行确定。

粒子群优化算法流程图：



图1. 13粒子群优化算法流程图

1. 初始化：

我们将粒子划分为：

个时段（低峰、平峰、高峰）；个相位，且每个方向-时段组合包含三个变量（左转、直行、右转的比例）

则一个粒子的纬度为36维。

初始化粒子的位置：

粒子的每 3 个连续变量表示某个方向+时段下的三个转向比例，且满足：

 且  （2）

所有粒子即可表示为：

 （3）

其中， 表示第时段第个方向上的左转车辆比例；表示第时段第个方向上的右转车辆比例；表示第时段第个方向上的直行车辆比例，其中，。

每个粒子的初始位置随机生成，即随机生成一组。

速度初始化为0，此时的最优解均为初始的位置本身。

1. 适应度函数：

设表示第时段第个相位上第种情况的估计总车流量。其中，。表示为第T时段上第相位的实际总车流量，因此，左转、右转和直行的车流量可以表示为：

1. 左转车流量：

 （4）

1. 右转车流量：

 （5）

1. 直行车流量：

 （6）

因此，我们可以确立最终的适应度函数：

 （7）

1. 迭代优化：

 （8）

 （9）

其中，表示第个粒子在时的位置，表示第个粒子在时的速度，为第个粒子在历史上找到的个体最优解，为所有粒子中找到的全局最优解，为惯性权重，控制粒子上一时刻速度对当前速度的影响，为个体学习率，衡量粒子对自身经验的信任程度，为社会学习率，衡量粒子群对群体最优经验的信任程度，为从之间随机生成的数，每次迭代都发生变化，用于增加搜索的随机性。

1. 收敛判断：

当达到最大迭代次数时，或者当粒子群算法得到左转和右转的比例均小于等于0或者大于等于0.5时停止。

最终我们通过Python代码求解，得到各相位上的不同情况的车辆比例如下：

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 时段 | 方向1左转比例 | 方向1右转比例 | 方向1直行比例 | 方向2左转比例 | 方向2右转比例 | 方向2直行比例 | 方向3左转比例 | 方向3右转比例 | 方向3直行比例 | 方向4左转比例 | 方向4右转比例 | 方向4直行比例 |
| 低峰 | 0.38 | 0.31 | 0.31 | 0.38 | 0.31 | 0.31 | 0.38 | 0.31 | 0.31 | 0.38 | 0.31 | 0.31 |
| 平峰 | 0.44 | 0.22 | 0.34 | 0.44 | 0.22 | 0.34 | 0.44 | 0.22 | 0.34 | 0.44 | 0.22 | 0.34 |
| 高峰 | 0.41 | 0.16 | 0.42 | 0.41 | 0.16 | 0.42 | 0.41 | 0.16 | 0.42 | 0.41 | 0.16 | 0.42 |

表1. 3节假日各时段各方向车辆左转、右转、直行比例表

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 时段 | 方向1左转比例 | 方向1右转比例 | 方向1直行比例 | 方向2左转比例 | 方向2右转比例 | 方向2直行比例 | 方向3左转比例 | 方向3右转比例 | 方向3直行比例 | 方向4左转比例 | 方向4右转比例 | 方向4直行比例 |
| 低峰 | 0.11 | 0.09 | 0.8 | 0.11 | 0.09 | 0.8 | 0.11 | 0.09 | 0.8 | 0.11 | 0.09 | 0.8 |
| 平峰 | 0.31 | 0.22 | 0.46 | 0.31 | 0.22 | 0.46 | 0.31 | 0.22 | 0.46 | 0.31 | 0.22 | 0.46 |
| 高峰 | 0.15 | 0.05 | 0.81 | 0.15 | 0.05 | 0.81 | 0.15 | 0.05 | 0.81 | 0.15 | 0.05 | 0.81 |

表1. 4工作日各时段各方向车辆左转、右转、直行比例表

因此，我们可以根据公式（4）、（5）、（6）进行计算工作日和节假日各时段各方向左转、右转、直行车流量，结果如下：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 时段 | 方向 | 左转车流量 | 右转车流量 | 直行车流量 |
| 低峰 | 1 | 1970 | 1612 | 14327 |
| 低峰 | 2 | 710 | 581 | 5165 |
| 低峰 | 3 | 1224 | 1002 | 8902 |
| 低峰 | 4 | 2399 | 1963 | 17452 |
| 平峰 | 1 | 11293 | 8015 | 16758 |
| 平峰 | 2 | 3440 | 2442 | 5105 |
| 平峰 | 3 | 8649 | 6133 | 12823 |
| 平峰 | 4 | 13390 | 9502 | 19869 |
| 高峰 | 1 | 22206 | 7402 | 119911 |
| 高峰 | 2 | 6890 | 2297 | 37203 |
| 高峰 | 3 | 16638 | 5546 | 89848 |
| 高峰 | 4 | 25220 | 8407 | 136190 |

表1. 5工作日各时段各方向左转、右转、直行车流量结果表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 时段 | 方向 | 左转车流量 | 右转车流量 | 直行车流量 |
| 低峰 | 1 | 4375.32 | 3569.34 | 3569.34 |
| 低峰 | 2 | 1941.04 | 1583.48 | 1583.48 |
| 低峰 | 3 | 2974.26 | 2426.37 | 2426.37 |
| 低峰 | 4 | 5095.8 | 4157.1 | 4157.1 |
| 平峰 | 1 | 8147.48 | 4073.74 | 6295.78 |
| 平峰 | 2 | 2785.64 | 1392.82 | 2152.54 |
| 平峰 | 3 | 4466.88 | 2233.44 | 3451.68 |
| 平峰 | 4 | 7318.52 | 3659.26 | 5655.22 |
| 高峰 | 1 | 48861.34 | 19067.84 | 50053.08 |
| 高峰 | 2 | 19005.55 | 7416.8 | 19469.1 |
| 高峰 | 3 | 32499.47 | 12682.72 | 33292.14 |
| 高峰 | 4 | 48406.65 | 18890.4 | 49587.3 |

表1. 6节假日各时段各方向左转、右转、直行车流量结果表

## 5.2问题二的模型建立与求解

### 5.2.1 数据准备

# 模型分析与检验

# 模型的评价、改进与推广

# 参考文献

# 附录：