**基于整数规划和神经网络的高速公路车辆轨迹数据分析**

**摘要**

近年来，随着我国高速公路交通流量大幅增加，为克服传统的全人工及半自动收费中存在的种种弊端，以不断提升高速公路通行效率和服务水平，电子收费系统(ETC)应运而生。该系统依托高速公路一定规模区域内密集、连续布设的门架系统，结合高速公路收费站等其他必要的数据资源，可以实现车辆行驶轨迹检测。

针对问题一，我们基于附件2作出了时间、门架和车辆通行数量三者关系的三维分布图，根据得到的图像，描述车流量的时空分布特征，也得到了维修门架的编号和维修的时间。

针对问题二，我们对附件2数据再次分析，发现的确存在门架漏失对过路车辆记录的情况。通过检索门架缺失记录，我们得到了各门架漏失过路车辆记录总数，进而可以求出各门架的漏失记录率。再对门架漏失过路车辆类型进行分析，得出了容易漏失记录的车辆的特征。

针对问题三，我们首先将每日收费站总流量和每日路段总流量进行对比，再将单位时间内通过收费站的车辆数和沿某一路段通过的车辆数量进行一一比较分析，作出体现两者关系的图像，进而综合分析收费站流量与路段流量的关系。然后，我们基于K-means聚类分析对各收费站进行分类，建立了单目标非线性整数规划模型，在保证通行顺畅的前提下，确定了各站收费车道数以及应急收费系统（便携式收费机或移动电站收费车）的数量，实现费用的最省。

针对问题四，

关键词：K-means聚类分析；多目标整数规划；季节指数预测

# **问题重述**

电子收费系统(ETC)广泛应用于高速公路车辆行驶轨迹检测，该系统包含了车辆视觉识别、天线、边缘计算服务单元等关键设备，主要依赖于高速公路一定规模区域内密集、连续布设的门架系统，结合高速公路收费站等其他必要的数据资源，实现车辆行驶轨迹检测。

**1.1问题一：**基于2022年2月22日至2月27日六天的数据，描述车流量的时空分布特征。并依据分布特征指出在规定时间范围内，运维部门维修门架的编号和维修的时间。

1.2问题二：判断门架是否会漏失对过路车辆的记录，如果会，指出各门架的漏失记录率和容易漏失记录的车辆的特征。

1.3问题三：说明收费站流量与路段流量的关系，并以在保证通行顺畅的前提下达到费用的最省为目的，确定各站收费车道数以及应急收费系统的数量。

1.4问题四：预测14个路段未来24小时各自的车流量，并依据预测模块实时优化未来24小时各个路段的限速值，以最大限度地确保路网安全畅通。

# **问题分析**

## 问题一的分析

针对问题一，基于附件2中2022年2月22日至2月27日的六天时间内经过该高速路区间的全部车辆的轨迹信息，由于完全基于所给数据进行统计分析，且数据量庞大，所以我们先对附件2的30个轨迹表进行数据预处理，发现并处理了异常数据。然后作出时间、门架和车辆通行数量三者关系的三维分布图，根据得到的图像，描述车流量的时空分布特征。在该段时间内，运维部门对一些ETC门架进行了维修，门架维修期间，记录到通过该门架的车辆数据信息显然会大幅降低。根据这个特点，结合三维分布图，可以得到维修门架的编号和维修的时间。

## 问题二的分析

针对问题二，我们首先将问题1中发现的附件2数据异常情况进行分类处理，发现的确存在门架漏失对过路车辆记录的情况。通过检索门架缺失记录，我们得到各门架漏失过路车辆记录总数，进而可以求得各门架的漏失记录率。再对门架漏失过路车辆类型进行分析，不难得到容易漏失记录的车辆的特征。

## 问题三的分析

针对问题三，我们首先将每日收费站总流量和每日路段总流量进行对比，为了进一步得到各收费站与各路段的关系，再将单位时间内通过收费站的车辆数和沿某一路段通过的车辆数量进行一一比较分析，作出体现两者关系的图像，进而综合分析收费站流量与路段流量的关系。然后，我们基于K-means聚类分析对各收费站进行分类，然后建立了单目标非线性规划模型，以通行时间和总费用为目标函数，以各站收费车道数以及应急收费系统的数量，车辆经过收费车道用时及车辆经过应急收费系统用时为决策变量，进行整数线性规划，得到整数可行域，进而可以得到最优解，使得在保证通行顺畅的前提下达到费用的最省。

## 问题四的分析

针对问题四，由于每天各路段车流量大体具有周期性，我们建立了季节指数预测模型，对各路段未来24小时的车流量作出了预测。

# **模型假设**

1. 假设两相邻门架间车速异常对数据分析结果的影响可以忽略。
2. 假设不考虑车辆在进出记录时间段以外被门架漏失记录轨迹的情况。
3. 假设门架6在维修期间必然漏失的记录都修补为可以正常记录。
4. 假设所有车型通过ETC与MTC时间相同。
5. 假设车道造价与车道数成正比。

# **符号说明**

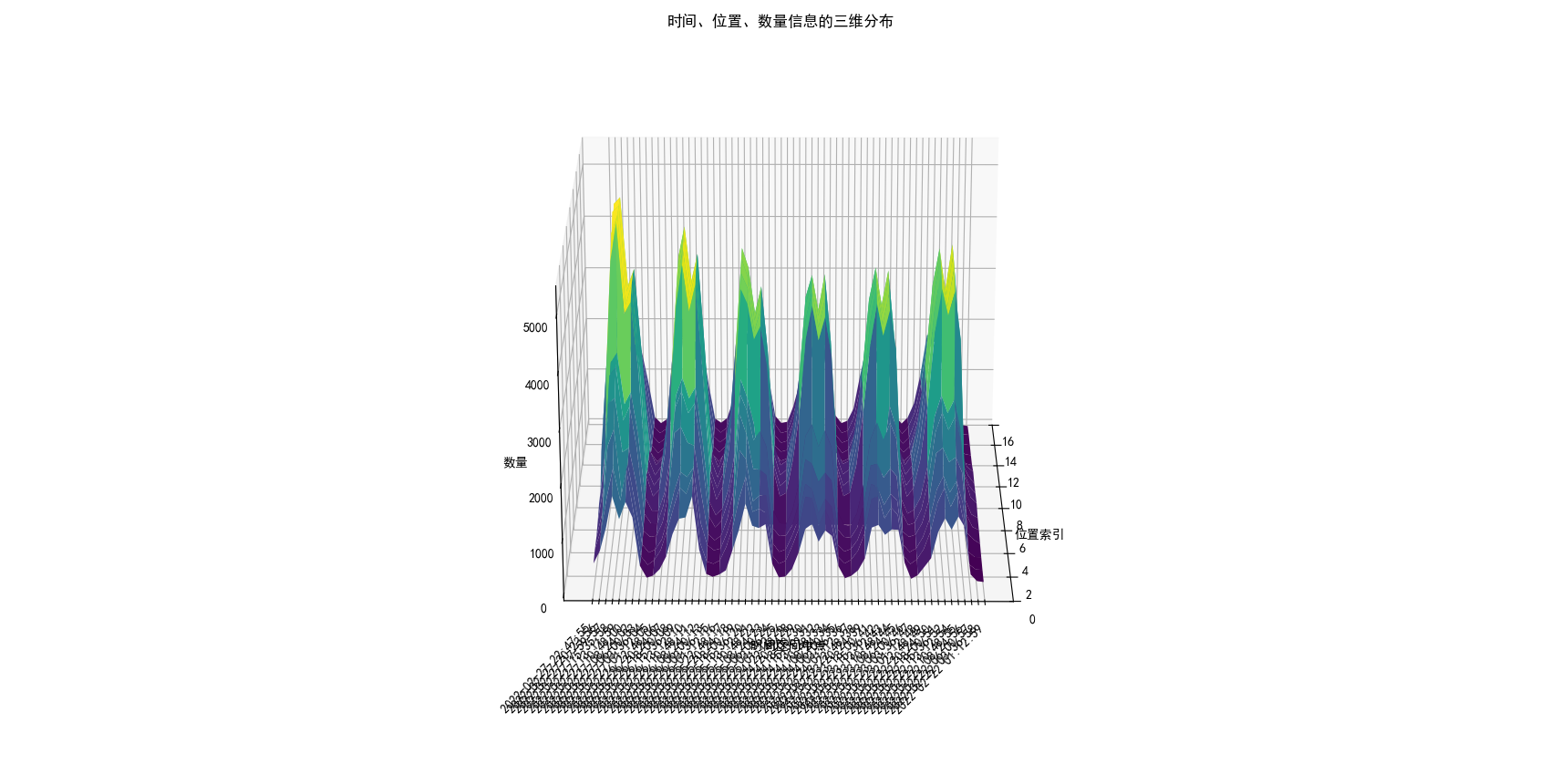
|  |  |
| --- | --- |
| 符号 | 说明 |
| latexmath | 第i号收费站修建车道及应急收费系统所用总费用 |
| latexmath | 第i号收费站修建的车道数 |
| latexmath | 第i号收费站修建的应急收费系统数 |
| latexmath | t时刻通过第i号收费站的车流量 |
| latexmath | 第i号收费站应急收费系统启动时间 |
| latexmath | 第i号收费站应急收费系统关闭时间 |
| latexmath | 第j个路段时间序列的全变动 |
| latexmath | 第j个路段长期趋势 |
| latexmath | 第j个路段季节变动 |
| latexmath | 第j个路段循环变动 |
| latexmath | 第j个路段不规则变动 |
| latexmath | 第j个路段t时刻的预测值 |
|  |  |
|  |  |
|  |  |

# **模型的建立与求解**

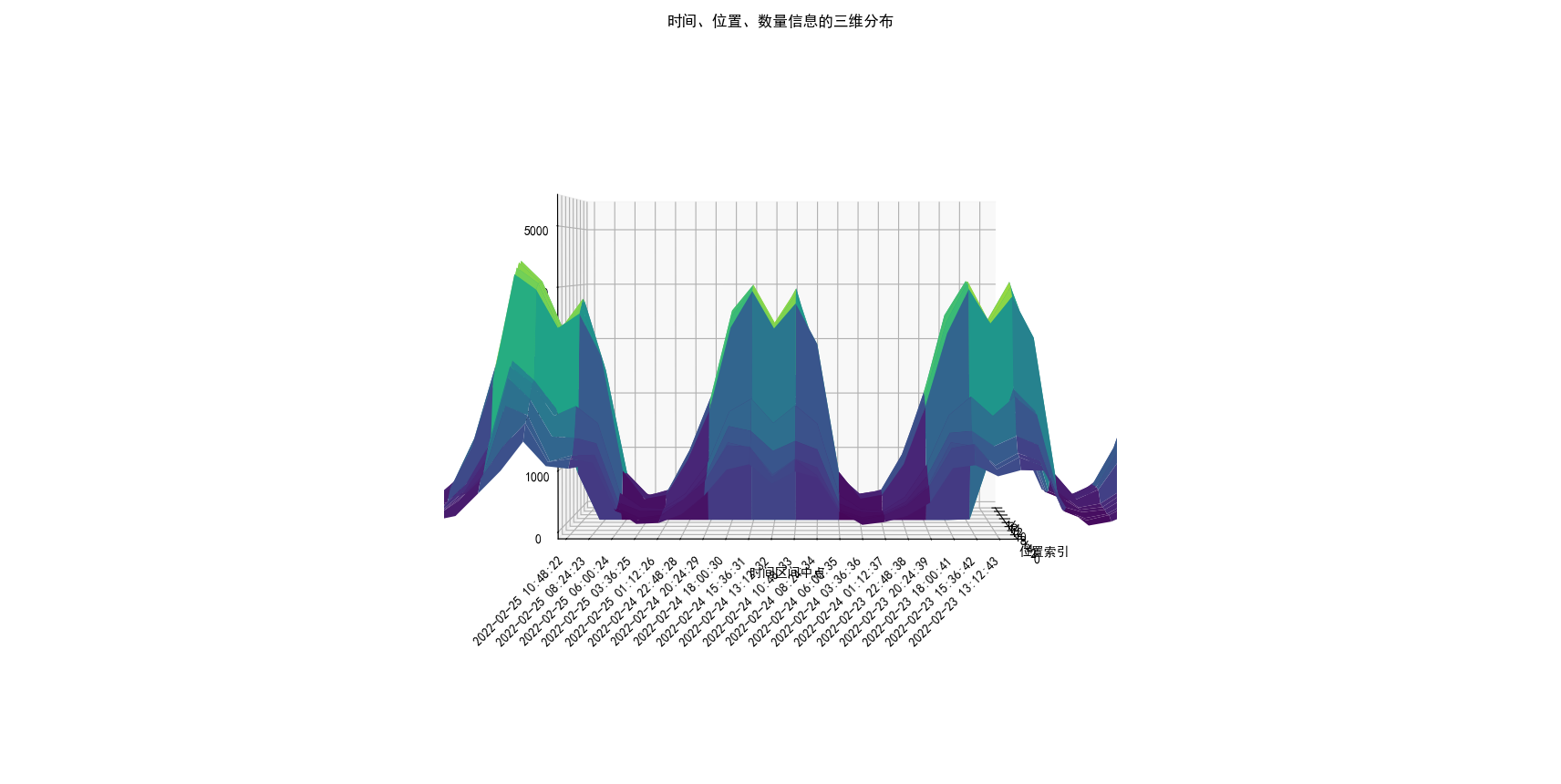
## 问题一模型的建立与求解

通过对附件2的数据检索，我们发现出现了极少数超出题目规定的分析时间段的轨迹记录以及出入站信息缺失的数据。由于30张轨迹表全体数据太过庞大，不妨以附件2轨迹表1为例，通过数据统计，总轨迹数为8646，而超出给定时间段的轨迹数为5条，出入站信息不完整的为5条，占总轨迹数的0.058%，对统计几乎没有影响。我们采用数据删除的方式，对符合题目条件的数据进行进一步处理。

在对附件2数据进一步观察中，我们发现极少数轨迹里出现门架记录漏失（在问题二中具体讨论）、通过门架编号顺序错乱、两相邻门架间车速计算异常的情况（误差时间短，不影响后续以小时为单位预测，在假设中已做处理）。针对门架编号顺序错乱，我们发现顺序颠倒的门架测量时间相差在3秒以内，粗略可以归为系统问题，直接进行顺序调换，归入两相邻门架间车速计算异常的情况，不影响分析。

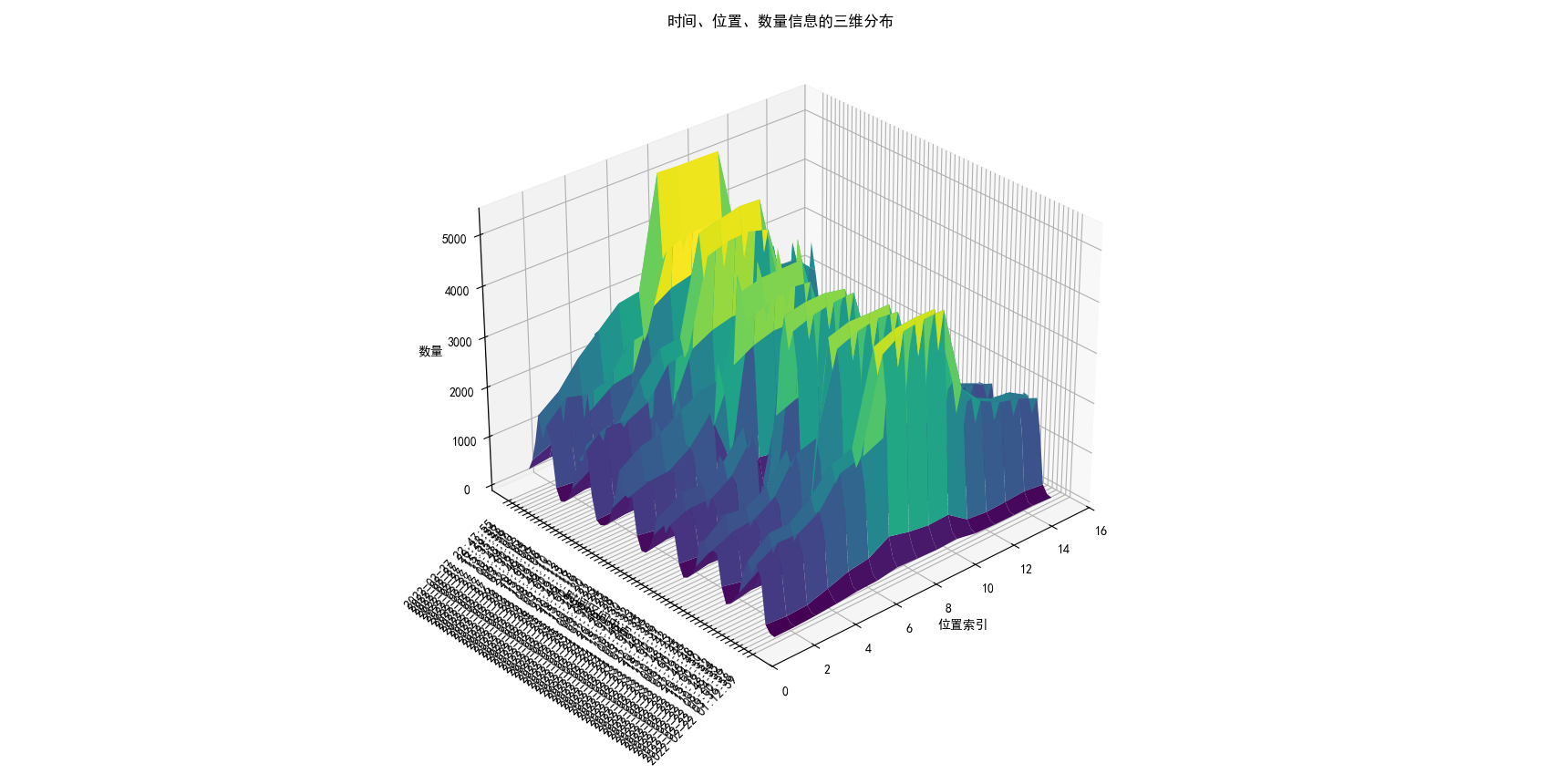


**图1 60等分时间区间各门架通行车辆数量（时间分布特征）**



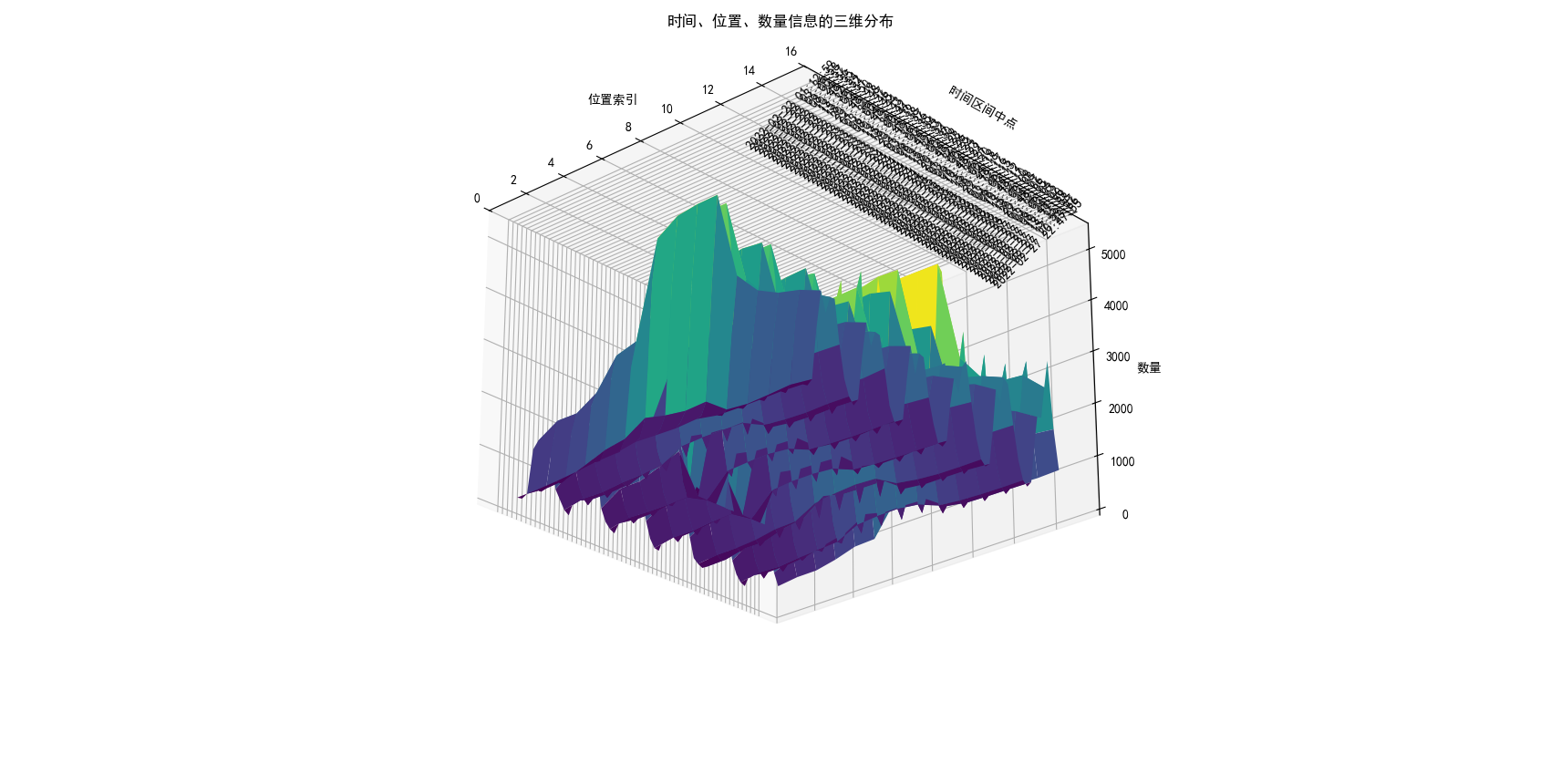
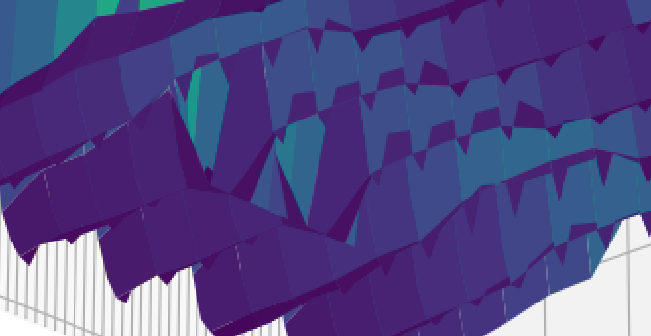
**图2 60等分时间区间各门架通行车辆数量局部放大图（时间分布特征）**

由图1可知，各门架通过车辆数量在时间分布上以24小时为一个最小周期进行循环增减。为得到更加精确的时间分布特征，我们对图1进行局部放大，得到了图2。由图2不难看出，在规定时间段内，每天各门架通过车辆数量在0时至10时逐渐增加，10时至13时会有小幅度下降，13时至15时会回升到10时水平甚至超越，15时至23时通过车辆数逐渐下降至当天0时水平。



**图3 60等分时间区间各门架通行车辆数量（空间分布特征）**

通过对图1进行方向变换得到图-3，我们可以以门架编号为主轴，观察车流量在空间上的分布特征。可以发现门架7、门架8、门架9、门架10相对来说通过的车辆最多。从图形上体现为靠近该路段中央，通过车辆较多；靠近该路段两端，通过车辆较少。

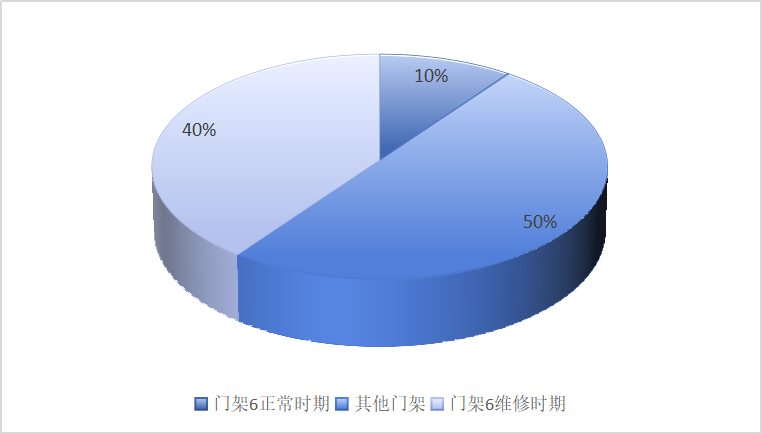
**图4 60等分时间区间门架维修特征图**

再次对图1进行方向变换得到图4，我们由图4可以观察门架维修情况。可以看出，门架6通过车辆数在2月24日为零，推出门架6为维修门架。为得到更加准确的维修时间，我们结合图2，可以发现门架6的维修时间大致位于2022年2月23日15时至2022年2月25日8时。

## 问题二模型的建立与求解

我们对附件2中门架记录通过车辆数据再次进行检索，以轨迹表1为例，发现共有39983条有效数据，1250条数据被漏失。由此，我们发现部分门架确实会漏失对过路车辆的记录。

我们按上述方法对剩余29个轨迹表漏失数据进行组成结构分析，综合得到如下结果：



**图5 附件2轨迹表漏失数据构成图（总）**

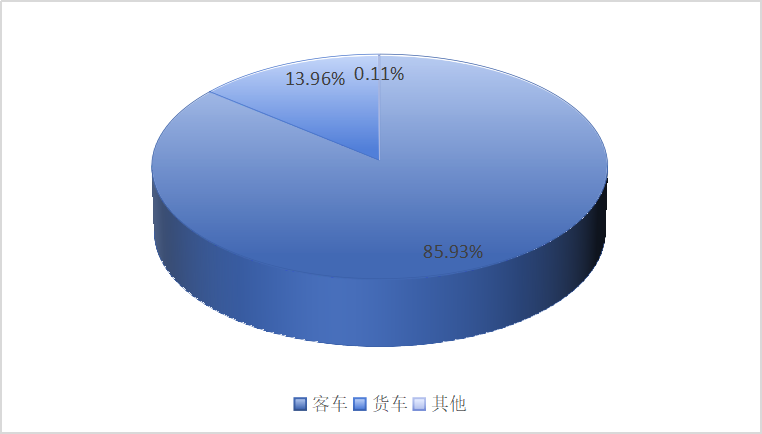
由问题一的解答与图5进行分析，门架6在2月23日15时至2月25日8时左右会进行维修。由此，我们作出假设，门架6在维修期间必然漏失的记录都修补为可以正常记录。同时对于漏过的门架事件，时间通过相邻事件的时间推断，以此保证各门架漏失记录率的准确性。

各门架的漏失记录率如表1所示。

**表1 各门架漏失记录率**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 门架编号 | 漏失纪录率 | 门架编号 | 漏失纪录率 | 门架编号 | 漏失纪录率 |
| 门架1 | 0.0000% | 门架6 | 4.7692% | 门架11 | 0.9739% |
| 门架2 | 1.4348% | 门架7 | 0.6715% | 门架12 | 4.9906% |
| 门架3 | 0.5650% | 门架8 | 1.0999% | 门架13 | 5.0973% |
| 门架4 | 0.2089% | 门架9 | 2.0154% | 门架14 | 1.2244% |
| 门架5 | 0.7224% | 门架10 | 1.4499% | 门架15 | 0.7224% |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

关于容易漏失记录的车辆特征的研究，我们作出了图6所示漏失车辆车型占比图。



**图6 漏失车辆车型占比**

由图6可知，在所有漏失记录的车型中，客车占比大幅度高于货车和其他车型。可以看出，相较于其他车型，客车更容易漏失记录。

## 问题三模型的建立与求解

针对分析收费站流量与路段流量的关系，先对每日的收费站总流量与路段总流量进行对比。此处我们选择轨迹表1与轨迹表2为例，分析收费站流量与路段流量的关系，得到结果如表2所示：

**表2 每日收费站流量和每日路段流量统计**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 2月22日 | 2月23日 | 2月24日 | 2月25日 | 2月26日 | 2月27日 |
| 收费站流量 | 3445 | 3075 | 2989 | 3432 | 3340 | 3252 |
| 路段流量 | 14702 | 12506 | 11859 | 13698 | 13743 | 13635 |

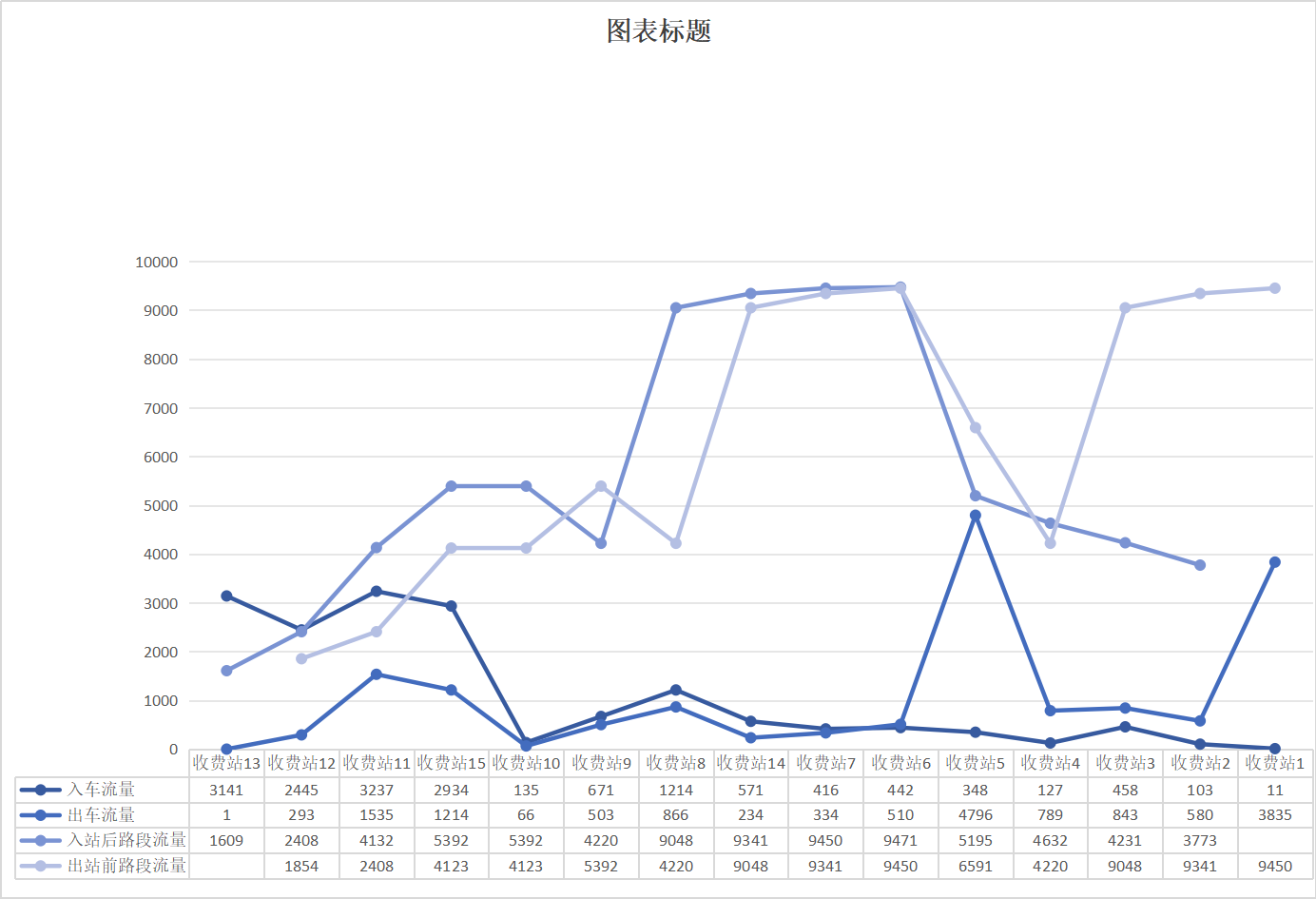
由表2可以看出，每日的收费站总流量与路段总流量成正相关。

基于上述结果，为了进一步得到各收费站与各路段的关系，我们根据附件1高速路段图中各收费站与各路段的相对位置关系，作出如表3对应关系，进行比较。

**表3 入出收费站与各路段比较关系对应表**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 入站收费站与路段 | | | | 出站收费站与路段 | | | |
| 收费站13 | 门架1 | 收费站14 | 门架8 | 收费站12 | 门架2 | 收费站7 | 门架8 |
| 收费站12 | 门架3 | 收费站7 | 门架9 | 收费站11 | 门架3 | 收费站6 | 门架9 |
| 收费站11 | 门架4 | 收费站6 | 门架10 | 收费站15 | 门架4 | 收费站5 | 门架11 |
| 收费站15 | 门架5 | 收费站5 | 门架12 | 收费站10 | 门架4 | 收费站4 | 门架12 |
| 收费站10 | 门架5 | 收费站4 | 门架13 | 收费站9 | 门架5 | 收费站3 | 门架13 |
| 收费站9 | 门架6 | 收费站3 | 门架14 | 收费站8 | 门架6 | 收费站2 | 门架14 |
| 收费站8 | 门架7 | 收费站2 | 门架15 | 收费站14 | 门架7 | 收费站1 | 门架15 |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |

根据表3中的对应关系，我们利用Excel软件分析处理出入收费站与各路段流量的关系，并作出图7所示图像。



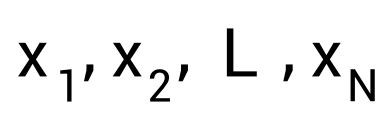
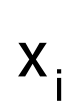
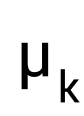
**图7 入出收费站与各路段流量折线图**

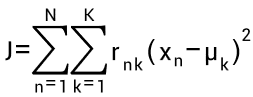
由图7可以看出，各收费站流量与对应路段流量大体成正比。收费站 8、11、12、13车流量较高，收费站14、3、15、5、1、7、 2车流量适中，收费站10、4、6、9车流量较低。收费站6、7、14、3、8对应路段流量较高，收费站5、2、4、9、10、15对应路段流量适中，收费站11、12、13对应路段流量较低。

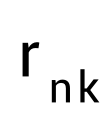
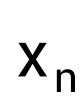
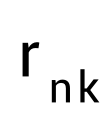
由于本文中涉及到的收费站车道数以及应急收费系统（便携式收费机或移动电站收费车）的数量均为整数，所以我们在此问选择基于K-means聚类分析的单目标线性整数规划模型。

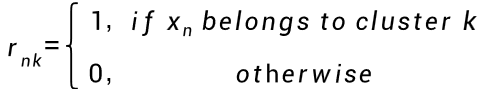
为便于计算，我们首先基于K-平均聚类算法对各收费站车流量进行聚类。

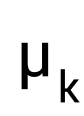
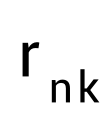
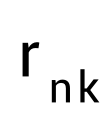
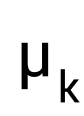
1. 平均聚类算法（K-means）是一种常用的无监督学习算法，用于将数据点分成不同的组或簇。其主要目标是将数据点分配到K个簇中，以使簇内对象具有较高相似度，而簇外相似度较低，其中相似度的计算根据一个簇中对象的平均值来进行。通过最小化数据点与其所属簇的中心点之间的距离的平方和来实现聚类。

我们假设存在N个数据点，每个数据点的维度为D。同时，假设要将数据点分成K个簇，每个簇有一个中心点，k=1,2,…,K。K均值的目标是最小化每个数据点与其所属簇的中心点之间的距离的平方和，即最小化如下损失函数：

 （1）

其中：是一个指示变量，当数据点属于第k个簇时为1，否则为0。通常情况下，可以表示为：

 （2）

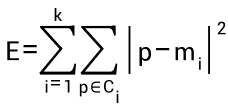
K均值算法的目标是找到最优的簇中心点以及相应的，使得损失函数J最小化。这通常通过迭代更新和来实现。K均值算法是一种迭代算法，其收敛性取决于初始簇中心点的选择以及数据的分布情况。

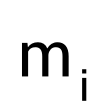
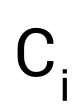
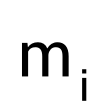
从n个数据对象中任意选择K个对象，作为初始聚类中心。

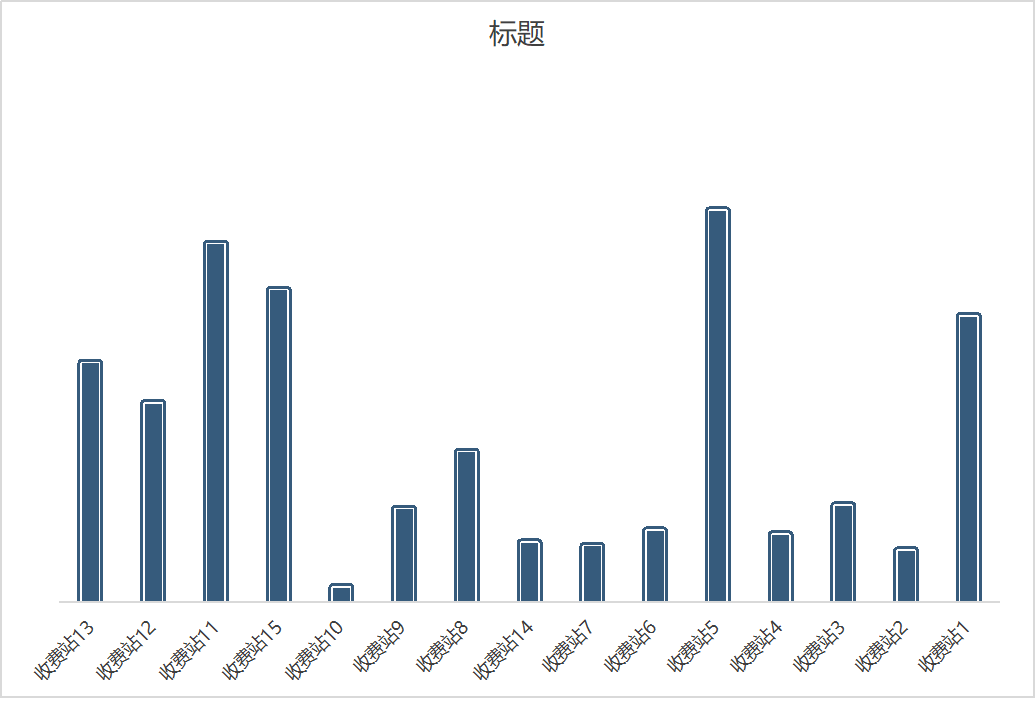
将每个数据点分配到离它最近的簇中心点所代表的簇。计算每个数据对象与各个簇中心点的距离，并由最小距离重新对相应对象进行划分，即将数据点分配到距离最近的簇中心点所代表的簇。

重新计算每个有变化的聚类的中心对象，即取该簇内所有数据点的平均值作为新的中心点。循环上述步骤，直到簇中心点不再发生变化或变化小于某个阈值。

通常选择均方差作为收敛准则函数：

 （3）

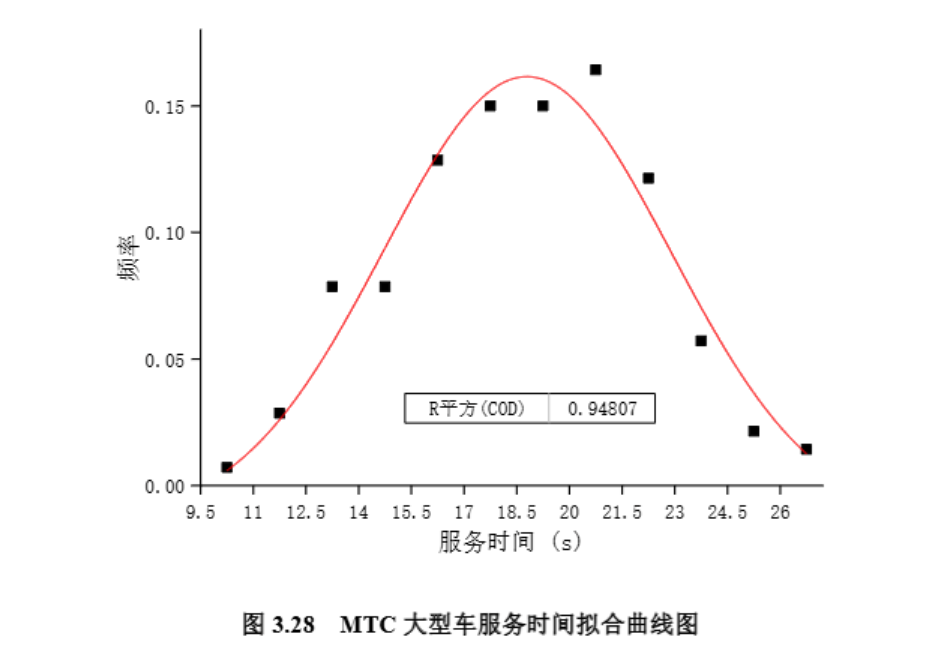
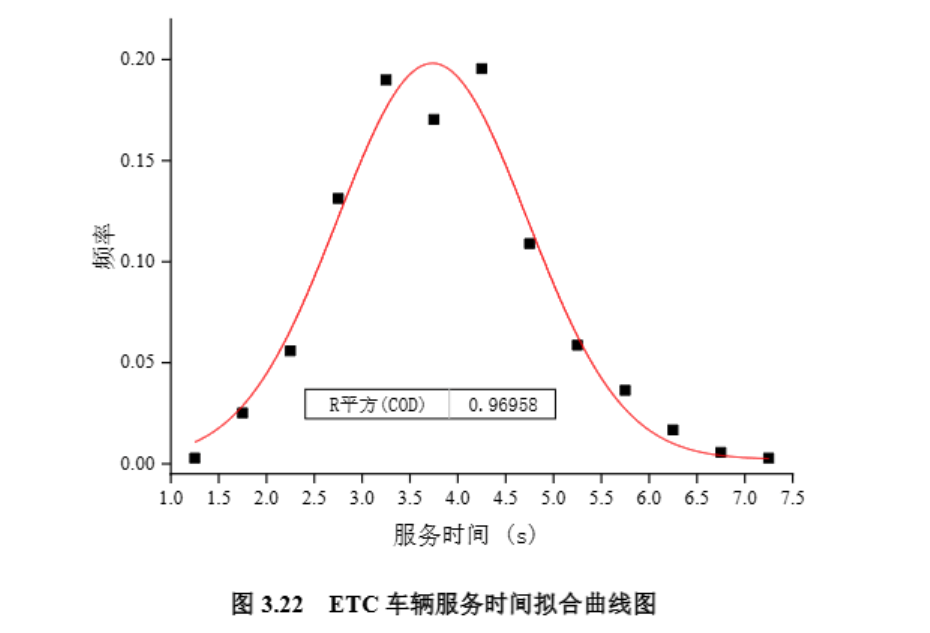
其中：E为数据库中所有对象的均方差之和；p为代表对象的空间中的一点；为聚类的均值。p和均是多维的。



**图8 各收费站出入车流量对比图**

我们通过聚类分析，依据图9所示各收费站出入车流量，将收费站分为3类：收费站5、11、15、1为第一类，收费站13、12、8为第二类，收费站9、3、6、4、14、7、2、10为第三类。

在具体计算中，我们认为所有车道都是ETC/MTC混合车道，不做具体区分。为了方便处理，我们通过大量数据得到ETC与MTC的车辆出入站比例为1：3，由此对两者通行效率进行加权，得到综合通行效率。



**图9 ETC通过车辆时间拟合曲线图 图10 MTC通过车辆时间拟合曲线图**

依据图7，图8的拟合曲线，我们设车辆ETC通过的时间为3.5s，MTC或补录通过时间为19s，代入模型中进行计算。

经过资料搜集与整理，我们粗略确定收费站车道每小时成本计算如下：

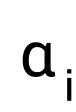
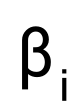
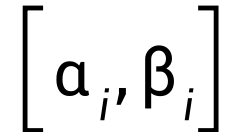
**单车道成本=单车道造价÷使用寿命（小时）≈25.8元**

单车道每小时能处理的车流量为：**3600÷（0.25×3.5+0.75×19）≈238辆。**

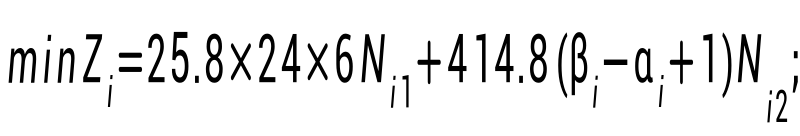
每套应急收费系统每小时成本计算如下：

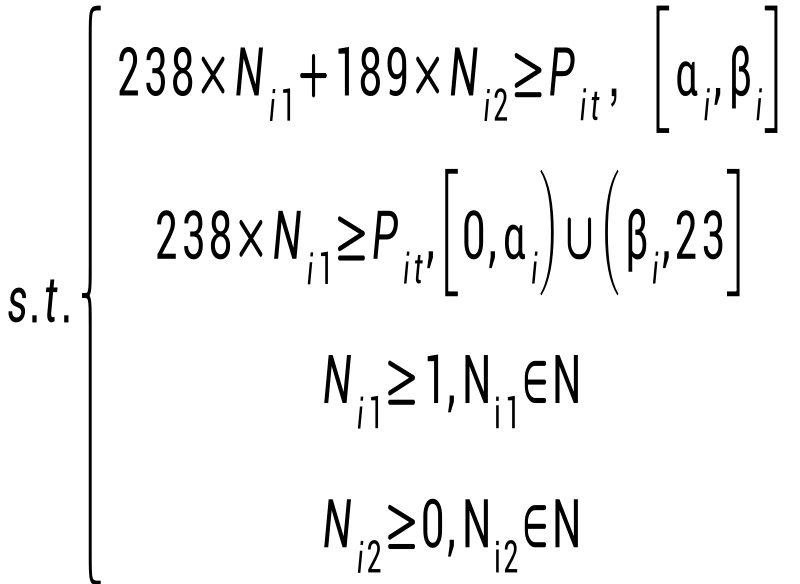
**应急收费系统成本=人员成本＋设备折旧＋能源消耗＋通讯费用＋维护修理+运输成本+保险成本≈414.8元**

每套应急收费系统每小时能处理的车流量为：**3600÷19≈189辆。**

通过查阅文献，我们了解到应急收费系统可以在车流量高峰时等效增加收费站车道数，即一个应急收费系统在车流量高峰期等价为一条MTC车道。由于应急收费系统并不是一天中每个小时都在工作，同时为了简化，我们以为第i个收费站应急收费系统每天开启的时间区间起点，以为时间区间终点，确定第i个收费站应急收费系统开启时间均为。

我们在对附件2进行数据处理时发现，车辆补录出站的时间未记录，又因不知道收费站具体位置，故我们采用经过的最后一个门架的时间代替其出站时间。

 （4）

 （5）

该模型接受某一收费站6天的逐小时车流量以及应急收费系统的开启时间区间，在此基础上尝试求得费用最优时，车道和应急收费系统的布局。在此基础上，再采取遍历所有的时间区间的方式，利用这一模型，寻找所有时间区间的最优费用中最优时的开启时间区间，从而得到针对该收费站的全局最优车道和应急收费系统的运营方案。

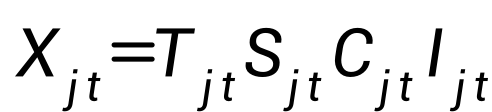
在对所有的收费站通行数据依次运用以上的方法后，我们得到了如下数据：

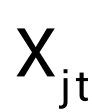
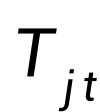
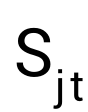
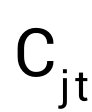
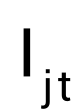
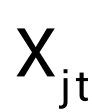
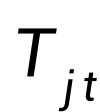
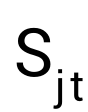
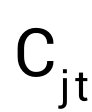
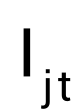
**表4 各站收费车道数以及应急收费系统数结果表**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 收费站 | 车道数 | 应急收费系统数 | 应急系统开启 | 应急系统关闭 | 总费用/元 |
| 收费站13 | 4 | 1 | 15:00 | 16:00 | 14860.8 |
| 收费站12 | 4 | 1 | 14:00 | 17:00 | 15690.4 |
| 收费站11 | 6 | 1 | 16:00 | 17:00 | 22291.2 |
| 收费站15 | 5 | 1 | 8:00 | 10:00 | 18990.8 |
| 收费站10 | 1 | 0 | － | － | 3715.2 |
| 收费站9 | 2 | 0 | － | － | 7430.4 |
| 收费站8 | 3 | 1 | 9:00 | 18:00 | 14464 |
| 收费站14 | 2 | 0 | － | － | 7430.4 |
| 收费站7 | 2 | 0 | － | － | 7430.4 |
| 收费站6 | 1 | 1 | 8:00 | 17:00 | 7033.6 |
| 收费站5 | 6 | 2 | 9:00 | 10:00 | 22291.2 |
| 收费站4 | 2 | 0 | － | － | 7430.4 |
| 收费站3 | 2 | 1 | 10:00 | 11:00 | 7430.4 |
| 收费站2 | 1 | 1 | 15:00 | 19:00 | 4959.6 |
| 收费站1 | 4 | 1 | 10:00 | 18:00 | 17764.4 |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

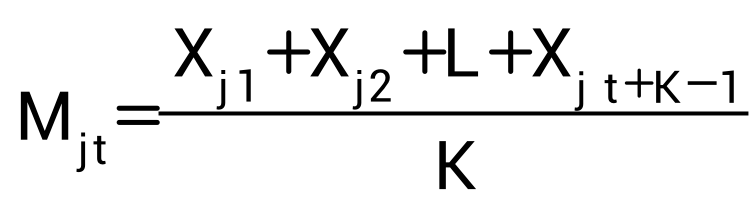
## 问题四模型的建立与求解

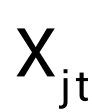
由问题一的解答可知，各路段车流量以一天为周期大致进行周期性变化，因此我们建立了季节指数预测模型，以预测各路段未来24小时的车流量。

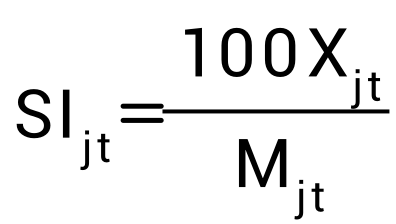
 （8）

式中为第j个路段时间序列的全变动，为第j个路段长期趋势，为第j个路段季节变动，为第j个路段循环变动，为第j个路段不规则变动。其中和用绝对数表示，，与用相对数（百分数）表示。

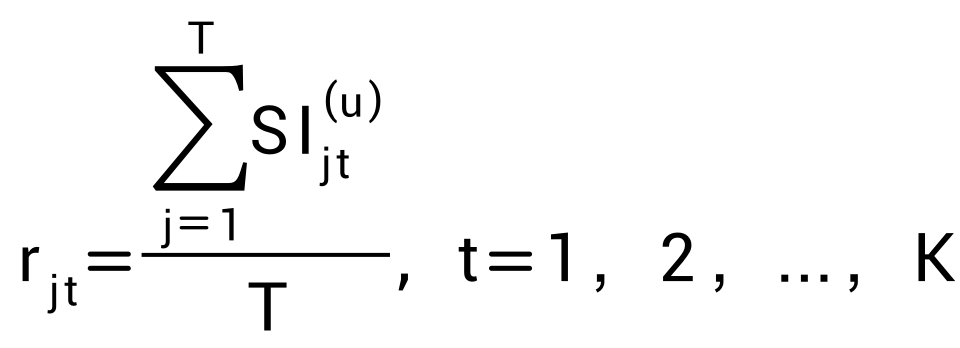
先对季节性S与随机性I求解：

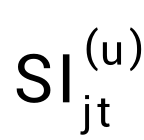
 （9）

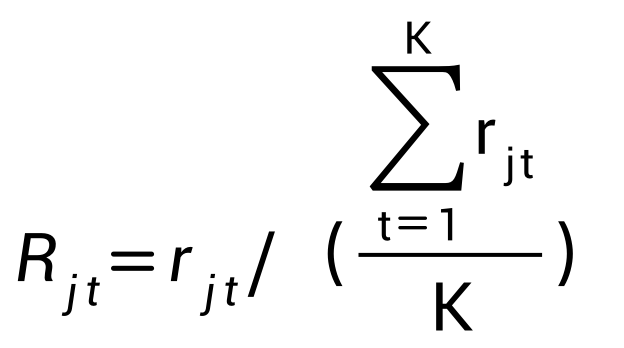
K为一个周期内的数据量，为第j个路段t时刻数据的观察值。

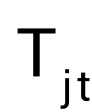
 （10）

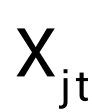
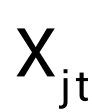
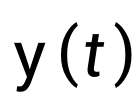
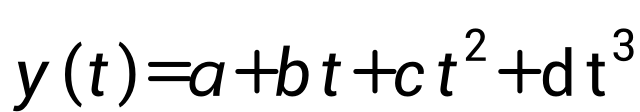
然后计算季节指数latexmath，计算各个周期数据点的平均值，相当从多个周期寻找一个周期内数据规律。

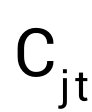
 （11）

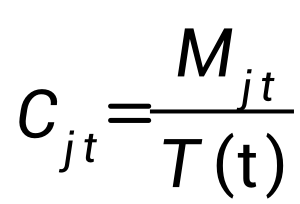
T为周期数量，为第j个路段第u个周期内第t个数据，设R为r修正后的数据，即为季节性和随机性：

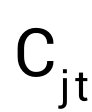
 （12）

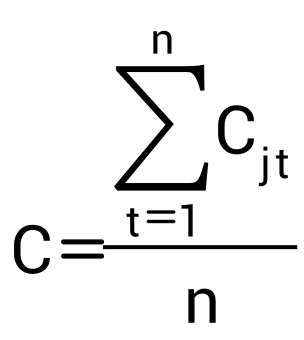
长期趋势的求解：

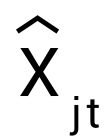
对数据进行拟合，根据整体趋势观察与t的整体关系，使用二次函数拟合，用表示拟合关系。可设，最后用最小二乘法求解出系数。

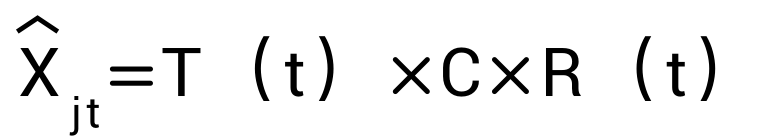
循环变动的求解：

 （13）

对求平均值可得到C：

 （14）

设第j个路段t时刻的预测值为，则：

 （15）

最终求得各路段未来24小时的车流量如图99所示：

图

# 模型的优缺点与改进

这部分不是必须的，没时间可以不写或者简写，当然写好了也是加分项。

## 模型的优点

列出一二三，一般百度搜一下你用到的模型或算法的优点就能找到。例如你用了prim算法，可以百度搜“prim算法优点”就能知道它适合解决边稠密顶点少的图。

## 模型的缺点

和优点一样，百度搜一下就有。要注意的是**缺点要写的避重就轻**，别把自己用的方法批判的太惨了。

## 模型的改进

一般是针对缺点来谈改进，可以说说还有什么方法能弥补你当前模型的缺点

# 参考文献

[1]常志恒. ETC推广对高速公路交通排放的减碳潜力评估[D].北京工业大学,2023.DOI:10.26935/d.cnki.gbjgu.2023.000274.

[2]苏帅杰. 基于推广ETC政策下的高速公路收费站通行能力和车道配置研究[D].长安大学,2021.DOI:10.26976/d.cnki.gchau.2020.001185.

[3]黄世彬.基于ETC数据的高速公路在途车辆路径预测研究[D].福建工程学院,2023.DOI:10.27865/d.cnki.gfgxy.2023.000240.

[4] 黄 志 敏 , 梁 承 东 . 基 于 K-means 聚 类 算 法 的 等 级 测 评 数 据 分 析 [J]. 电 子 质 量,2023,(12):40-44.

[5]孙晶.高速公路收费站车道配置与建设成本对比分析[J].黑龙江交通科技, 2019, 42(6):4.DOI:CNKI:SUN:HLJJ.0.2019-06-093.

附录