**基于XXXX模型的高速公路车辆轨迹数据分析问题研究**

摘要

（1）**标题**一般写成 基于XXXX模型的XXXX问题研究 就可以了。

（2）摘要是**最重要**的部分！要做到让评阅老师读完摘要，就能大体上知道论文写了什么。总共分为三部分：开头段、针对每小问各一段、结尾段。要在正文都写完后再来写摘要，高度总结概括正文内容。

（3）**关键词**一般写问题的关键词、模型名称、算法名称等，4~8个即可。

（4）由主要负责论文写作的队员写好摘要，其他两名队员当做评审来看，一定要反复阅读修改。

总体内容形式：

开头段：本文针对XXX问题，通过建立XXX模型，实现对XXX的求解。

针对问题一，考虑XXX，…，建立XXX**模型**/利用XXX**方法**，…**求解出**XXX。

针对问题二，考虑XXX，…，建立XXX模型/利用XXX方法，…求解出XXX。

针对问题三，考虑XXX，…，建立XXX模型/利用XXX方法，…求解出XXX。

结尾：总结下全文，继而分析下你用的模型的优点，自夸一下。

关键词：XXX，XXX，XXX

# **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1，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2，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

# **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。

对于小白来说，往往写不出来太好的假设。实在没办法的话，可以写一些正确而无用的话来凑点内容。（当然能写出有用的话最好）

常见的假设：

1. **对问题影响非常小的因素**，例如求解一天中影子长度变化，地球公转对影长的影响远远小于自转，那么假设“忽略一天中地球公转对影长的影响”；
2. **将模型中的某些无法掌握的变量设为常数**，从而简化模型。例如关于有些问题涉及到价格，可以“假设XXX的定价在XXX过程中始终不变”。（一般针对这种假设需要进行灵敏度分析）
3. **一些小概率事件，或者说正确而无用的话**。例如关于物质浓度测量的题，假设“试纸没有过期、无破损”

# 符号说明

|  |  |
| --- | --- |
| 符号 | 说明 |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

正文当中一些重要的或者多次用到的变量都写在这。

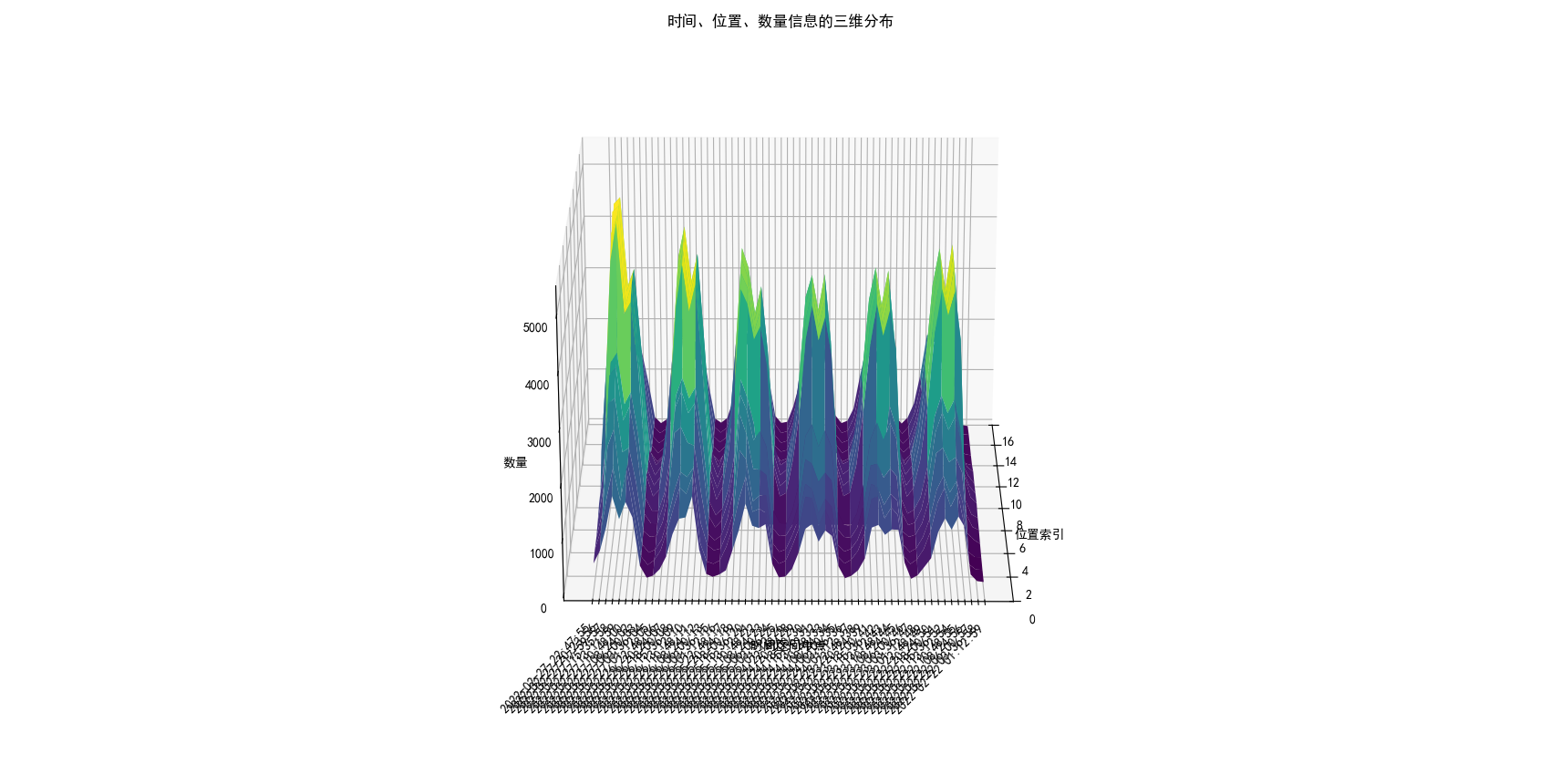
1. 只在某几步用到、一些临时变量就不用写了。
2. 说明要简略，不用详细讲述
3. 正文中第一次用到变量，还是需要解释的

# **模型的建立与求解**

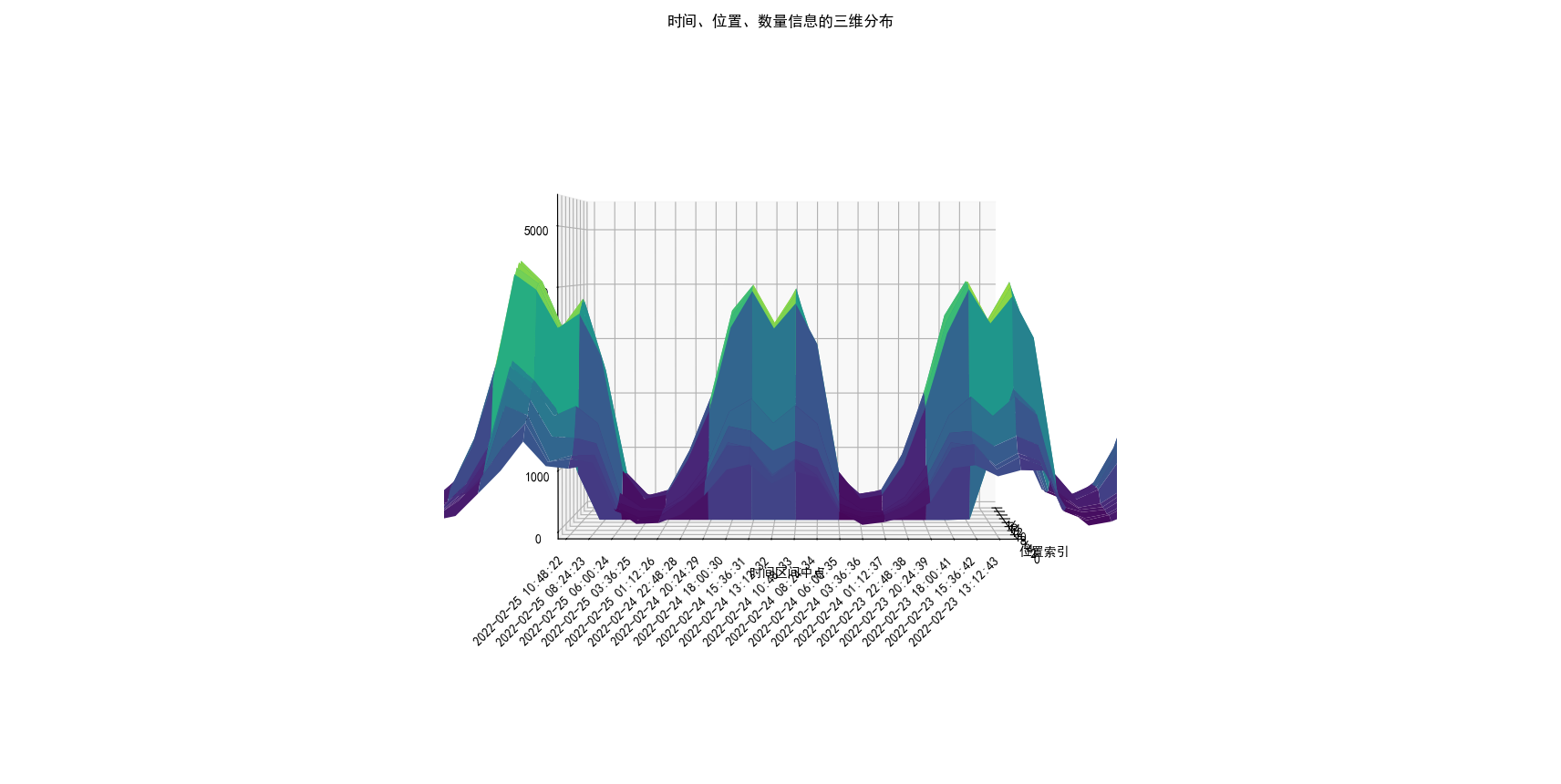
## 问题一模型的建立与求解

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。

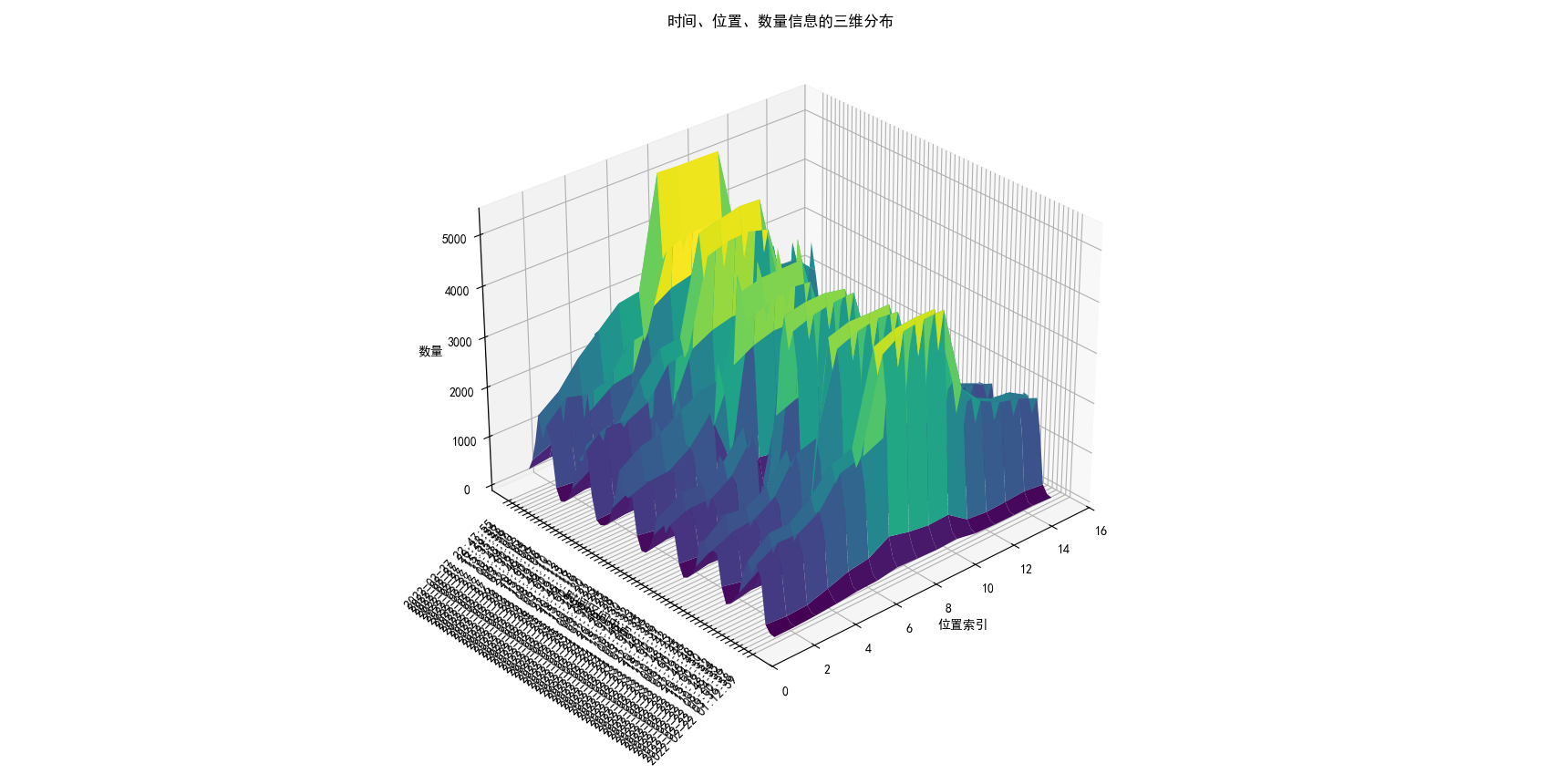


**图1 60等分时间区间各门架通行车辆数量（时间分布特征）**



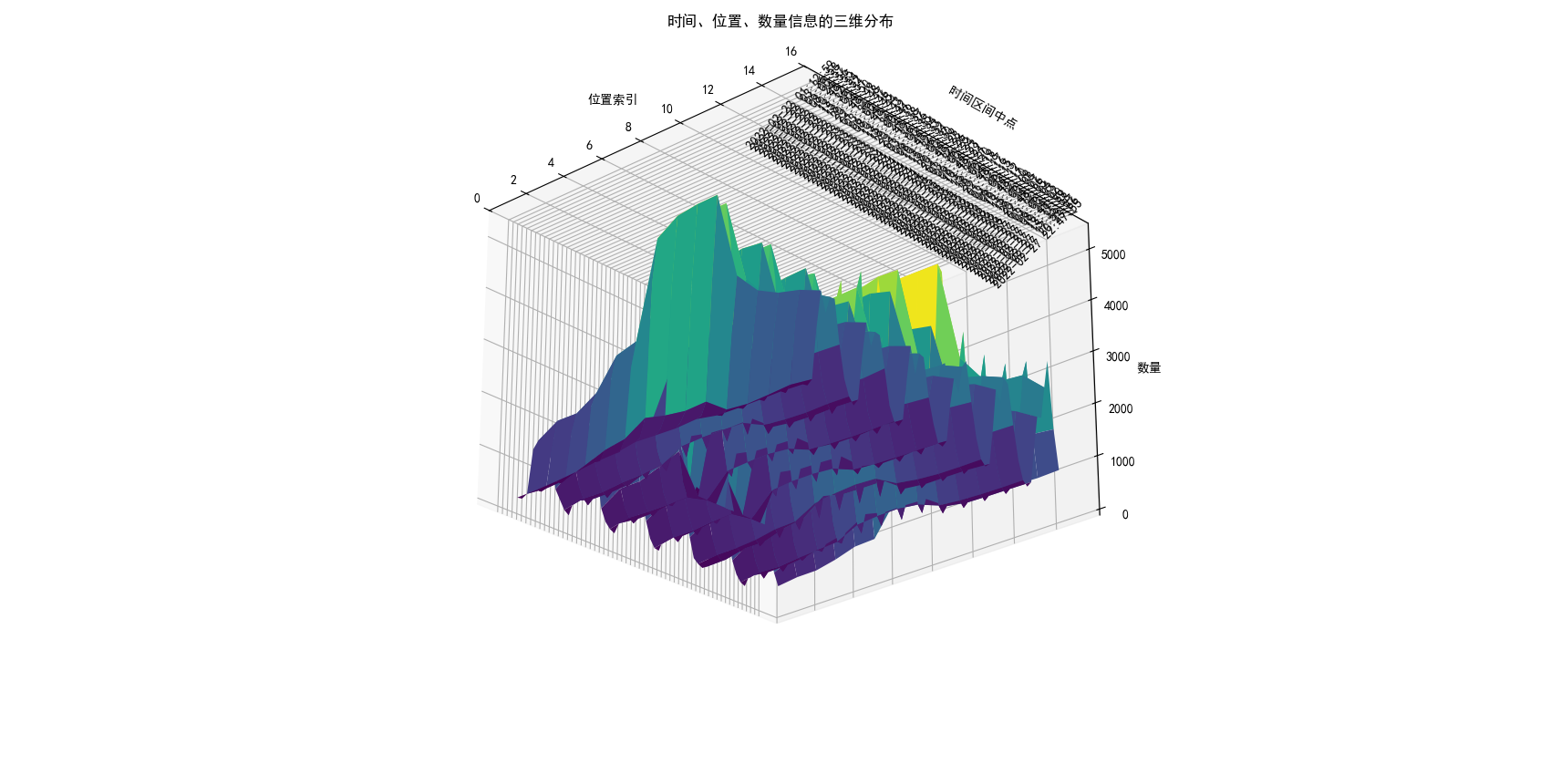
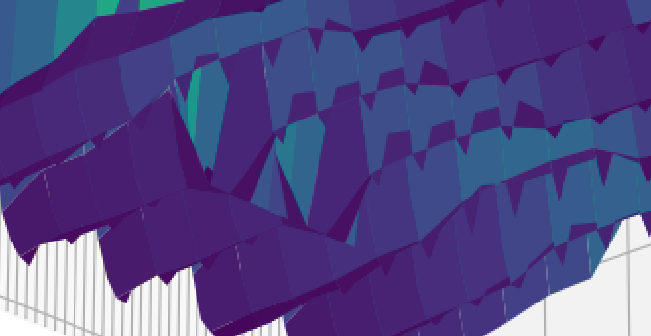
**图2 60等分时间区间各门架通行车辆数量局部放大图（时间分布特征）**

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。

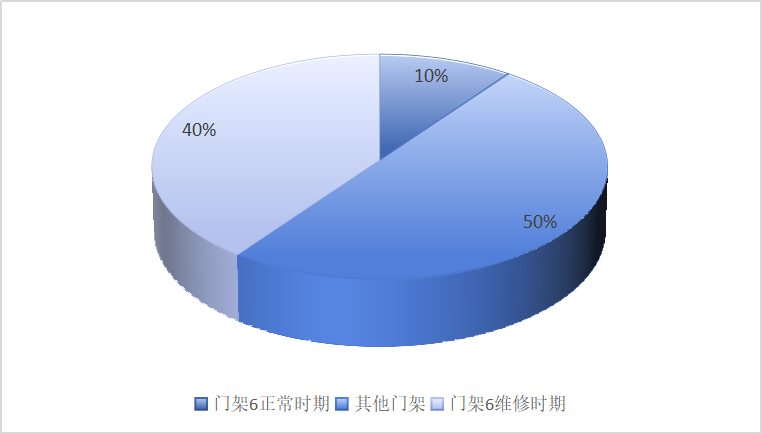
**图4 60等分时间区间门架维修特征图**

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通过车辆数在某些时刻为零，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：



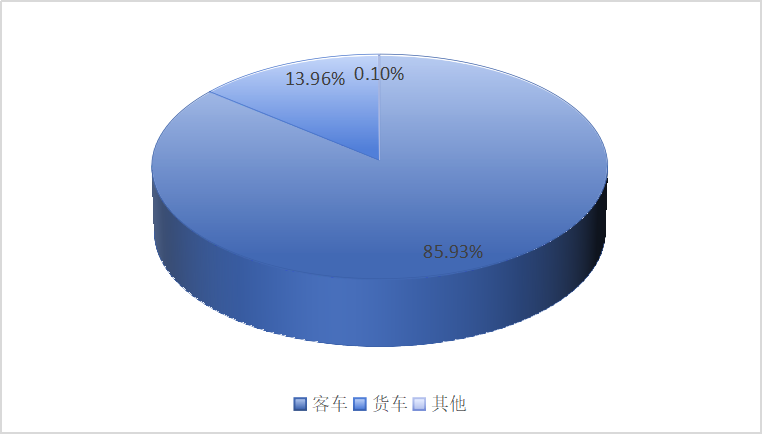
**图5 附件2轨迹表漏失数据构成图（总）**

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% |
|  |  |  |  |  |  |
|  |  |  |  |  |  |



**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

## XXX模型的改进

|  |  |
| --- | --- |
| 插入公式 | () |

## XXX结果分析

结果分析部分，需要**明确回答赛题的问题**。

如果赛题是**最优化问题**，一般需要进行灵敏度检验。

1. **什么是灵敏度检验？**在模型假设部分的举例：“假设XXX的定价在XXX过程中始终不变”，现实生活中物价是随时变化的，我们做这个“物价不变”的假设是为了简化问题。
2. 现在基于“物价不变”的最优解求出来了，但**现实生活中当物价真的变了的时候**，**求出的这个最优解还会是最优的么**？
3. 显然，**假如物价变动非常小**，涨了或跌了几分钱、几毛钱，那一般最优解还是最优的
4. 但如果涨了两倍，或者便宜了一半，那现在求的最优解可能就不是“最优”了。我们要解决的就是，到底物价变化到多少时，最优解会变了呢？

总之，最优性检验就是把最初为了省事而设的“常数”视作“变量”，分析这个“变量”对我们求解结果的影响有多大。

注意！虽然有些资料里强调灵敏度检验多么重要，但**这不是每篇论文都必须有的**！灵敏度检验只适用于预测类和最优化类的问题，也就是建模过程中有一些现实中本来是变量、但在模型中被假设为常数的情况。

对于小白来说，连最优化模型都不一定能建好求解出来，更不可能真懂最优性检验了。所以这一部分只是**锦上添花**，**一定要先把模型建好求解完在考虑最优性检验等操作**，**不要舍本逐末！**

# 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。

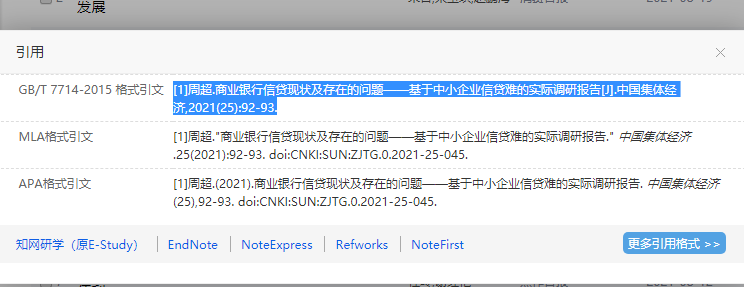
## 模型的改进

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

在知网查文献，检索结果的左右边的标志，点开就有标准的引用文

把这段话复制到论文里就行了。



[1]周超.商业银行信贷现状及存在的问题——基于中小企业信贷难的实际调研报告[J].中国集体经济,2021(25):92-93.

附录

附录把用到的代码放进来就可以了。这部分并不重要。

注意：

* 1. 用网上现成的代码求解问题完全可以，但**绝对不能把网上的代码直接复制粘贴到论文附录里**！！！曾经有过被省里推到国奖结果被查出代码是抄的而被取消获奖并全网通报的先例。
  2. 只需要把**代码里的变量名改一改**就行，比如原代码用的a，全都改成b就行了。
  3. 除了代码，还可以放一些比较冗长的表格。在正文当中求解出某个结果，需要几十行的表格展示，那么可以在正文相应的地方写一句：“具体数据见附录”，然后把表格放在附录就行
  4. 更加冗长的数据等放在支撑材料里。