

第四届“泰迪杯” 全国数据挖掘挑战赛

优
秀
作
品

作品名称：基于时间序列模型的铁路旅客流量预测

荣获奖项：特等并获企业冠名奖

作品单位：中国计量大学

作品成员：李正 沈烨平 朱慧慧

指导老师：王义康

基于时间序列模型的铁路旅客流量预测

摘要: 本文旨在研究铁路旅客流量预测问题,通过对数据的挖掘,多维度的分析客流规律,建立客流规律预测模型并优化车辆配置及车辆停靠方案。

本文通过运用 eclipse, mySQL 等软件建立了列车到站离站信息数据库,包括列车表,站点表,到站离站表和区间表等常用表。用 matlab 结合 mySQL 对旅客梯形密度表中的数据从车次、时段、车站和区间四个角度进行了分类处理。综合考虑各车次的客座率,客流流入区间波动情况,每日的客流高峰时段,站点等级和客流流向,多维度的分析客流规律。针对站点,运用改造的 RFM 模型对各个站点分区;针对区间,从三个角度——ZD111-01→ZD190-01 的本地客流(管内上行)、ZD190-01→ZD111-01(下行)的本地客流和 ZD111-01→ZD190-01 的外地客流分析客流流向趋势。

其次,本文建立考虑季节特征的 SARIMAX 模型对未来客流进行了预测,但是节假日效应对预测误差影响明显。基于数据的分析,我们发现节假日的预测误差在期间以三角函数波形波动,我们采用傅立叶函数系对节假日误差段进行三阶拟合,拟合优度都在 75%以上。同时结合预测值与实际值线性变化的性质,对前一年节日误差段进行当前年份的校正,并用 15 年春节数据对 16 年春节校正并检验,效果显著。这足以证明这种**基于傅立叶拟合的节假日误差修正技术**适合于消除 SARIMAX 模型所不能规避的节假日效应。

同时考虑天气的影响,发现天气虽然会影响客流量,但是对预测误差影响不明显,我们认为这是由于 SARIMAX 模型已经在客流量数据变化的角度解释了这种现象。

基于这两种情况的分析,给出了未来两周各搭乘区间客流量的预测。

最后,建立多阶段决策的调度模型。考虑到管内车客座率分配不均会导致高峰期列车资源的严重浪费的情况,兼顾模型求解的复杂性,建立了以**日客座率各车次平均分配**为目标的 0-1 规划模型。用 lingo 分支定界得到各车次不同日期的停靠站点的局部最优分配方案,同时对比节假日(高峰期)和工作日(低谷期)分配方案,得到常驻站点和可变站点的设置方案。此外,结合客流规律的分析,采用贪婪算法,对编组方案进行了优化设计。具体如下:取消 D11 和 D18 班次的 ZD250 站点;D16 增开 ZD062 站点;D05, D06, D07 和 D16 增开 ZD190-02 站点;D05 车次由 16 组编组改成 8 组编组。

关键词: 改造的 RFM 模型 SARIMAX 预测模型 傅里叶拟合 0-1 规划 贪婪算法

The thesis title

Abstract: This paper aims to study the prediction of railway passenger traffic. A new model for predicting railway passenger flow based on data mining and multidimensional analysis of passenger data has been proposed. Furthermore, guiding by the established system, the railway train operation plan has been optimized.

Current passenger data are analyzed by trips, time, stations and section by using Java, eclipse and mySQL. Firstly, the rule of passenger volume is multi-dimensional evaluated by passenger load factors of all train number, fluctuations of passenger flow into the section, the daily flow peak hours of passengers, site rating and traffic flow, analysis of traffic laws. RFM model was used for optimization of the stations, while the passenger volume for each section was diagnosed by three special cases, e.g. local passenger volume between ZD111-01 and ZD190-01, local passenger volume between ZD190-01 and ZD111-01 and foreign passenger volume between ZD111-01 and ZD190-01.

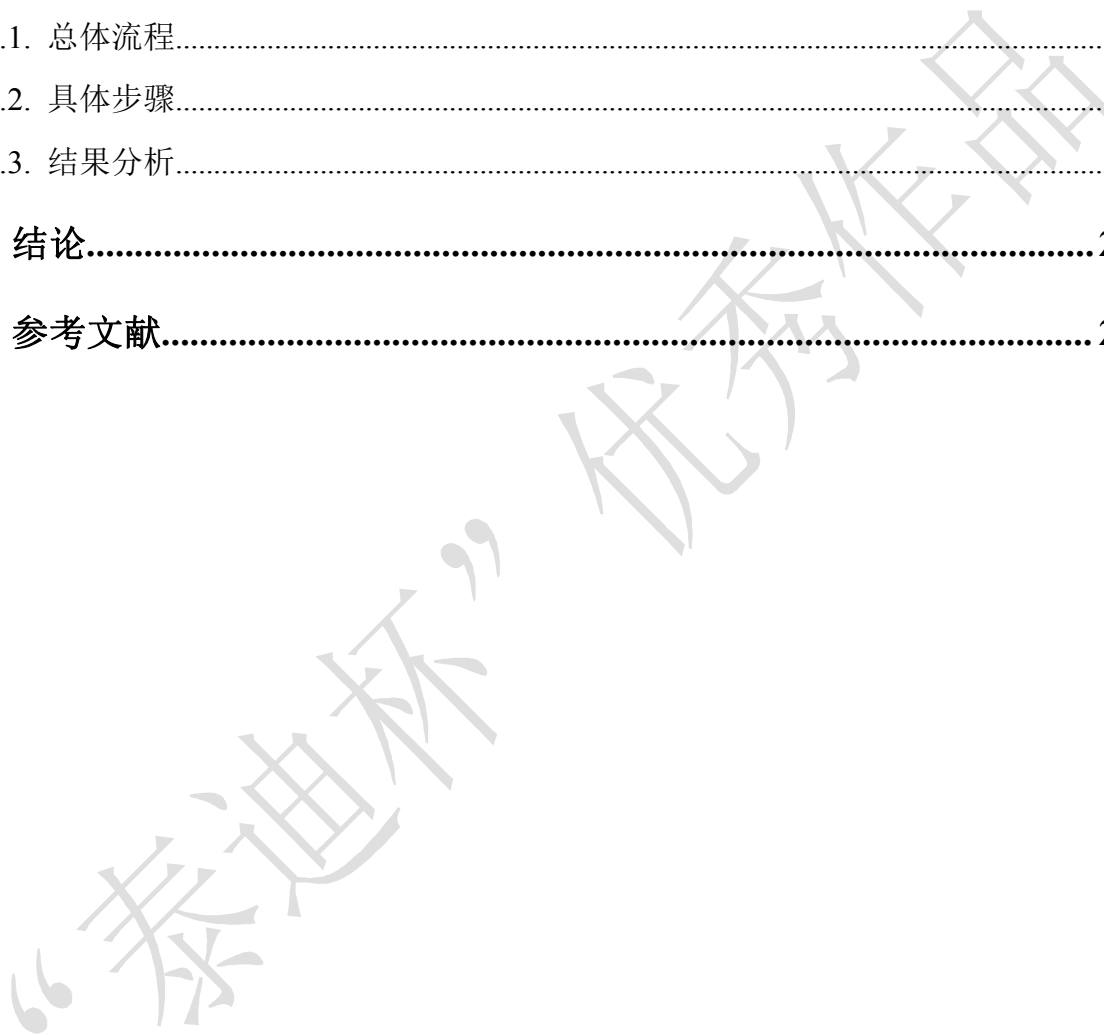
Secondly, prediction model of SARIMAX was set up to forecast the future passenger flow. Besides, compared to the effect of weather, the influence of holidays was found to be more obvious, which is modified by using Fourier fit in the model, while the effect of weather is ignored in the model.

Lastly, multi-stage scheduling decision model was built to optimize the railway train operation plan. However, in consideration of the complexity of the model solution, the model was simplified to the 0-1 programming problem of mean distribution of daily load factor trips. Then, combined with rule of passenger volume, the optimization of the railway train operation plan was achieved by using greedy algorithm. Particularly, D11 and D18 in ZD250 is cancelled; station ZD062 is added to D16; station ZD190-02 is added to D05, D06, D07 and D16; D05 is changed from 16-grouped to 8-grouped.

Key words: RFM model, SARIMAX model, Fourier fit, 0-1 programming, greedy algorithm

目录

1. 挖掘目标.....	4
2. 分析方法与过程.....	4
2.1. 总体流程.....	4
2.2. 具体步骤.....	5
2.3. 结果分析.....	25
3. 结论.....	26
4. 参考文献.....	26



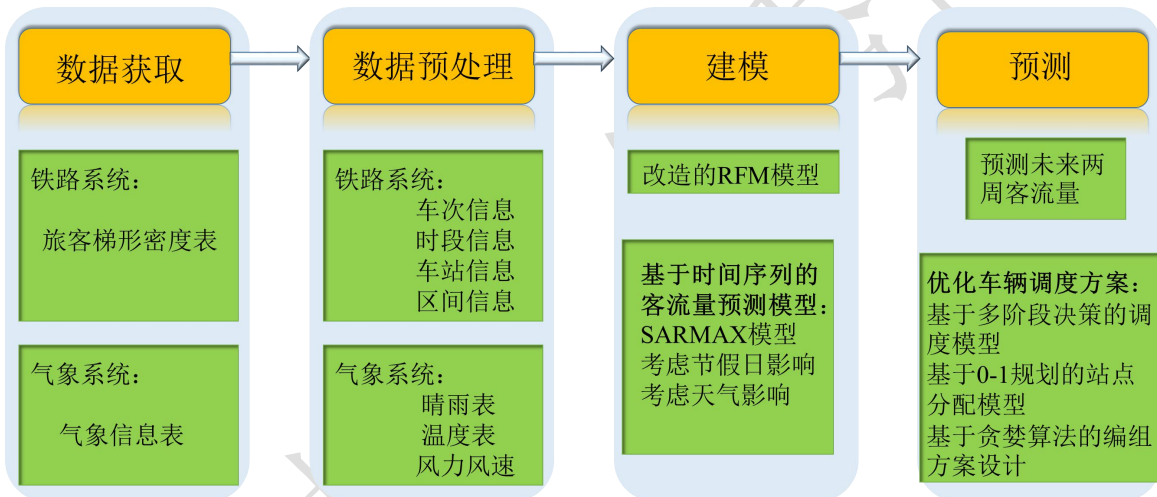
1. 挖掘目标

本次建模利用铁路系统中积累下来的海量真实数据，采用数据挖掘技术，达到以下三个目标：

- (1)按车次、时段(小时)、车站、区间(两个车站之间)等条件分析客流规律。
- (2)考虑节假日和天气等相关因素的影响，从管内和管外客流两个角度，建立时间序列预测模型，并预测未来两周的客流量。
- (3)针对 D02-D19 车次的列车，优化设计车辆配置及车站停靠方案。

2. 分析方法与过程

2.1. 总体流程



本用例主要包括如下步骤：

步骤一：铁路系统数据和气象系统数据的分类采集是本次数据挖掘分析的第一步。运用 Java, eclipse 和 mySQL 等软件，对旅客列车梯形密度信息进行车次信息，时段信息，车站信息和区间信息的挖掘。同时对所统计的地区气象信息进行天气状况，温度状况和风速信息的挖掘。

步骤二：数据预处理，考虑到直接采集的部分数据并不能直接拿来进行分析，通过运用相关的模型和数学分析，有针对性的处理数据，得出需要的信息。

步骤三：数据分析：针对车次信息，在考虑管内车还是管外车的情况下分析该车在区间内的客座率。对管外车进一步分析客流流入区间的情况。针对时段信息，分析一天内站点的客流变化情况。针对车站信息，运用改进的 RFM 模型来衡量站点价值。针对区间信息，观测两站点区间内客流量，分析乘客的流动规律。

步骤四：根据步骤三中数据分析得出的客流规律，将列车分为管内车和管外车，综合考虑天气和节假日的影响因素，建立 SARMAX 时间序列预测客流量模型，并进一步的运用该模型来预测未来两周的客流规律。

步骤五：根据前面分析的客流规律，针对 D02-D19 车次的列车，设计基于 0-1 规划

的站点分配问题，并通过采用贪婪算法来调节车辆的编组方案。

2.2. 具体步骤

结合总体流程图，对每一步骤做详细说明。

步骤一：数据的分类采集：

根据题目中所给的数据，包括旅客列车梯形密度表和车站所属地区气象信息表。为从车次，时段，车站和区间等角度分析客流规律，我们整理做出如下的 5 张分类信息表：

表 1 为通过附件 3 整理得到的站点信息，如下所示：

表 1：站点信息表

站点名	里程	地区
ZD111-01	0	ZD111
ZD111-02	3	ZD111
ZD311	49	ZD111
ZD326	110	ZD326
ZD192	153	ZD250
ZD022	186	ZD250
ZD250	210	ZD250
ZD062	295	ZD250
ZD120	320	ZD190
ZD121	330	ZD190
ZD143	341	ZD190
ZD370	362	ZD190
ZD190-02	378	ZD190
ZD190-01	393	ZD190

表 2 为车次信息表，以日为单位分析每趟车的客流规律。考虑到本地车只运输本地客流，而长途车不仅包括本地客流也包括外地客流，因此首先将车次分成两类，本地车和长途车，用 0(本地车)和 1(长途车)标记。为分析列车在所给区间内客流情况，通过客座率这一指标来衡量。客座率的计算公式如下：

$$\frac{\sum \text{旅客运输量} \times \text{运输距离}}{\text{座位数量} \times \text{运输总公里数}}$$

此外，由于 ZD190 和 ZD111 对应两个经济发达的大型旅游城市，有较多的外地客流，可通过记录区间整体客流流出或流入的数量来分析出本地客流和外地客流在该区间的流动规律。记客流流入为+，流出为-。

具体的操作方法可利用统计区间内各站点的上下车人数，考虑到每趟车的终点站均为 ZD190，所以没有客流流出管内，我们可通过分析外地客流流入的多少来分析游客的规律。下面给出 2015 年 1 月 1 日数据的一部分，如表 2 所示，一年的数据统计表见附件 2。

表 2：车次表

2015.01.01			
列车名	管内客座率	本地车/长途车	客流流入
K11	0.390	1	+638
G01	0.412	1	+689
D02	0.956	0	0
Z02	0.314	1	+487

T01 0.466 1 +638

表 3 为时段表，分析每天各个时间段的站点上下车人数，以统计该时段的客流密度。每隔 1 个小时记录一次上下车人数。通过该表，我们可清晰的观察到每天各个时间段内站点的人流高峰期，以便及时做好高峰时段的车站增加安保和维持秩序的服务工作。下面给出 2015 年 1 月 1 日 ZD111-02 站点各个时间段内上下车人数，各个站点一年的数据统计表见附件 2。

表 3：时段表

2015.01.01				
时间	站点	上车人数	下车人数	
0:00-6:00	ZD111-02	83	227	
6:00-7:00	ZD111-02	266	0	
7:00-8:00	ZD111-02	338	9	
8:00-9:00	ZD111-02	237	230	
9:00-10:00	ZD111-02	237	230	
10:00-11:00	ZD111-02	470	157	
11:00-12:00	ZD111-02	321	133	
12:00-13:00	ZD111-02	321	133	
13:00-14:00	ZD111-02	224	200	
14:00-15:00	ZD111-02	172	239	
15:00-16:00	ZD111-02	172	239	
16:00-17:00	ZD111-02	348	152	
17:00-18:00	ZD111-02	348	152	
18:00-19:00	ZD111-02	117	93	
19:00-20:00	ZD111-02	198	131	
20:00-21:00	ZD111-02	198	131	
21:00-22:00	ZD111-02	5	249	
22:00-23:00	ZD111-02	5	249	
23:00-24:00	ZD111-02	0	0	

表 4 为车站表，记录各个站点一天上下车的人次，以便分析该站点人流量的多少。为衡量站点为大站还是小站，综合参考实际数据，我们人为规定年人流量超过 600 万的站点即为大站。为进一步分析站点信息，我们引入改造的 RFM 模型，综合考虑站点人流量和获取的利润，较为精确的衡量站点价值。改造的 RFM 模型具体表述为，通过一个站点的人流量的多少以及获取的利润这两项指标来描述该站点的价值状况。为此表 4 中进一步的根据票价将车次分成两类：高铁、动车(GD)和普快、特快、直达特快(KTZ)，并记录站点两类车的停靠车次及上车人数，来估算不同站点的利润高低。下面给出 2015 年 1 月 1 日各站点的信息，各站点一年的数据统计表见附件 2。

注：针对长途车到点的时间如果与始发时间不在一天的情况，在对某站点统计某天的车次时按该车到该站点的日期算。

表 4：车站表

2015.01.01				
站点名		上车	下车	停靠车次
ZD111-01	GD	11513	7101	48
	KTZ	0	0	0
ZD111-02	GD	0	0	0
	KTZ	0	0	0

ZD311	GD	572	883	4
	KTZ	0	0	0
ZD326	GD	3960	7820	48
	KTZ	0	0	0
ZD192	GD	795	2163	13
	KTZ	0	0	0
ZD022	GD	144	387	2
	KTZ	0	0	0
ZD250	GD	3899	9282	48
	KTZ	0	0	0
ZD062	GD	620	2019	9
	KTZ	0	0	0
ZD120	GD	0	0	0
	KTZ	0	0	0
ZD121	GD	19	162	1
	KTZ	0	0	0
ZD143	GD	0	0	0
	KTZ	0	0	0
ZD370	GD	0	0	0
	KTZ	0	0	0
ZD190-02	GD	497	1140	11
	KTZ	0	0	0
ZD190-01	GD	3761	10281	43
	KTZ	0	0	0

表 5 为区间表，用于分析区间内的人流量的大小，衡量各个区间内车内拥挤状况，以及客流的流向情况。这里值得注意的有三点：①区间分析时要考虑列车行驶的方向，如 A 站 → B 站的客流较多，我们不能一概而论为 A 站-B 站该区间的客流较多，需要将客流的流向考虑进来。②同一市区不同部分站点之间人员的流动较小，可忽略不计。这一近似具有实际意义，通常市内交通比较发达，人们通常不会选择火车这一交通工具。③本地车和长途车需要分开考虑，长途车承载旅客的客流较多，可能在旅游城市下车的人数较多，而本地车的客流流向可能就不具备这一趋势。下面给出 2015 年 1 月 1 日各区间的信息，各区间一年的数据统计表见附件 2。

表 5：区间表

2015.01.01	
站点 → 站点	人流量
ZD111-01 → ZD111-02	本地车：6106 长途车：3761
ZD111-02 → ZD311	本地车：6106 长途车：27134
ZD311 → ZD326	本地车：5950 长途车：26921
ZD326 → ZD192	本地车：5359 长途车：22684
ZD192 → ZD022	本地车：5041 长途车：21529
ZD022 → ZD250	本地车：4871

	长途车: 21456
ZD250 → ZD062	本地车: 3779
	长途车: 14466
ZD062 → ZD120	本地车: 1916
	长途车: 7642
ZD120 → ZD121	本地车: 565
	长途车: 4435
ZD121 → ZD143	本地车: 565
	长途车: 4183
ZD143 → ZD370	本地车: 565
	长途车: 3464
ZD370 → ZD190-02	本地车: 565
	长途车: 2889
ZD190-02 → ZD190-01	本地车: 355
	长途车: 1986
ZD190-01 → ZD190-02	本地车: 3761
	长途车: 0
ZD190-02 → ZD370	本地车: 4160
	长途车: 0
ZD370 → ZD143	本地车: 4160
	长途车: 0
ZD143 → ZD121	本地车: 4160
	长途车: 0
ZD121 → ZD120	本地车: 4160
	长途车: 0
ZD120 → ZD062	本地车: 4160
	长途车: 0
ZD062 → ZD250	本地车: 4099
	长途车: 0
ZD250 → ZD022	本地车: 4491
	长途车: 0
ZD022 → ZD192	本地车: 4491
	长途车: 0
ZD192 → ZD326	本地车: 4417
	长途车: 0
ZD326 → ZD311	本地车: 4031
	长途车: 0
ZD311 → ZD111-02	本地车: 1272
	长途车: 0
ZD111-02 → ZD111-01	本地车: 163
	长途车: 0

步骤二：数据的整理分析：

根据前面几张表格中整理的数 据，我们分别从车次，时段，站点和区间等角度分析

客流规律:

①车次

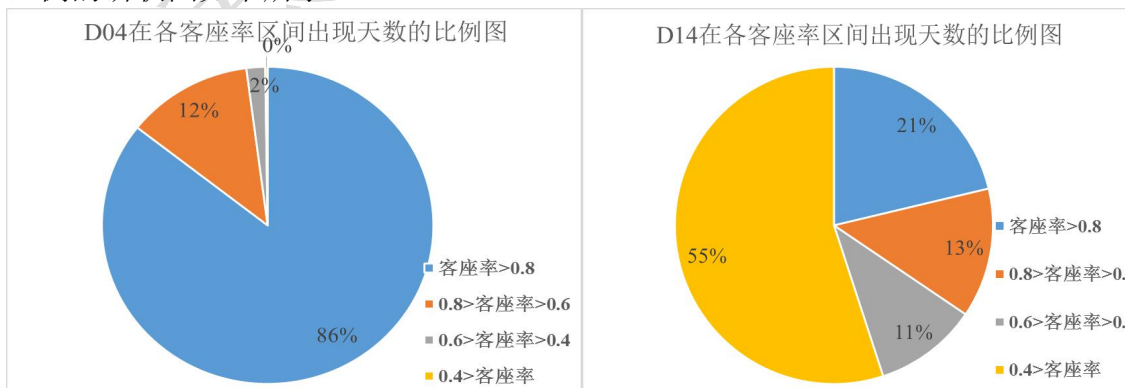
1)首先, 根据只在所给区间内行驶还是跨区间将车次分为本地车和长途车。不难发现本地车为 D02-D20、D49-D51, 其他车次的车均为长途车。其次, 通过对不同车次的车一年内管内各个站点所经过的次数的统计, 我们发现两点比较有特色的结论: ①一年之中部分车次的车作为临时增开的班次, 增开班次一般有以下几种情况: 节假日客流较多时, 如元旦。值得注意的是, 这并不是车次增开的主要原因。通常增开班次都是在 6 月份或第二年的 1 月份, 很大可能是经过半年或一年的铁路运营后, 对相关线路调整分析后决定增开的车次。②针对一趟车, 管内有些站点是常驻站点, 还有些是可变站点。我们有选择性的挑选几个典型车次的可变站点设置的时段, 如下表所示:

表 6: 典型车次可变站点设置的时段表

车次	可变站点及设置的时段
D06	ZD192: 1-5 月份 ZD121: 6 月份以后增开站点
D08	ZD190-02: 5 月份以后增开站点 ZD190-01: 1-5 月份
D11	ZD192: 5 月份以后增开站点
D14	ZD311: 5-6 月份
G04	ZD022: 1-5 月份
G05	ZD121: 2016 年增开站点
G11	ZD190-02: 2016 年增开站点
G26	ZD311: 6 月份以后增开站点
K01	ZD370: 1-2 月份
K15	ZD190-02: 2 月下旬增开站点
K18	ZD311: 1-5 月份

从上表我们可看出, 站点增开或不开大多数是经过几个月的运行或一年的运行后总结客流规律之后铁路局做的相应的调整。从中可看出, ZD190 市的客流量较大, 几个车次增开的站点都设在 ZD190 市, ZD250 市的客流量相对较小, 取消了部分该市的站点。

2)针对非临时增开的班次, 根据表 2 中统计的一年的管内客座率信息, 可分析出该车次的载客率的高低。其中有几个车次的客座率比较典型, 例如: 动车 D04 一年的整体的客座率都非常高, 动车 D14 和 D16 一年的客座率都相对较低。高铁各个车次的客座率普遍比较高, 其中最为突出的是 G06, G09。普快普遍客座率不高, 相对来说 K07 客座率太低的天数相对较少。针对几个比较典型的车次, 做出其在各客座率区间出现天数比例的饼状图如下所示:



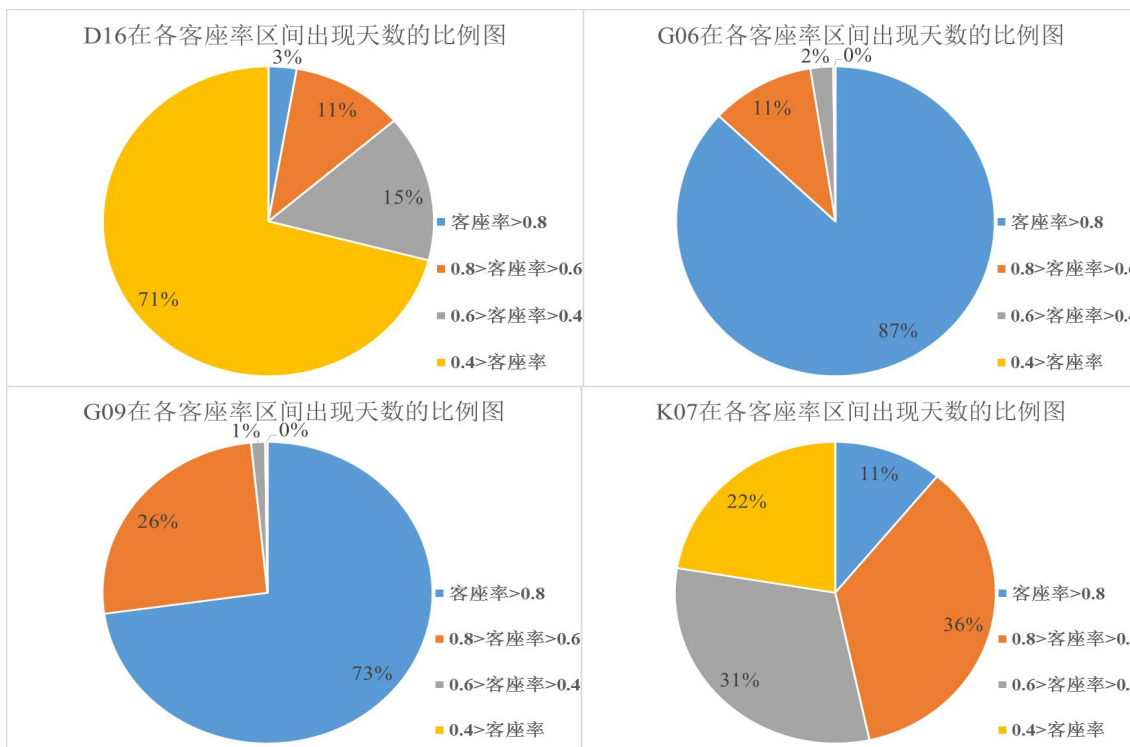


图 1: 典型的车次在各客座率区间出现天数的比例图

3)此外，我们引入客座率比的概念，即：

$$\text{客座率比} = \frac{\text{管内客座率}}{\text{总客座率}}$$

当客座率比>1 时，表明管内线路载客率高于总体线路，这意味着该管内线路相较整体线路而言设置是较优的。经分析，高铁大部分车次的该条线路设计是较优的，从其在管内很高的客座率也可看出，但是其中也有些车次客座率普遍低于总客座率，比如 G02，G03，G07，G08。分析其原因，针对 G02，G03，由于这两趟车经过的站点较多，管内部分已接近终点站，所以相应的客座率会减小。对 G07，G08 而言，可能是发车时间安排不恰当，有调整的空间。

4)针对长途车，可通过表 2 分析客流流入区间的信息，我们做出如下的流入区间的客流随时间变化的图：

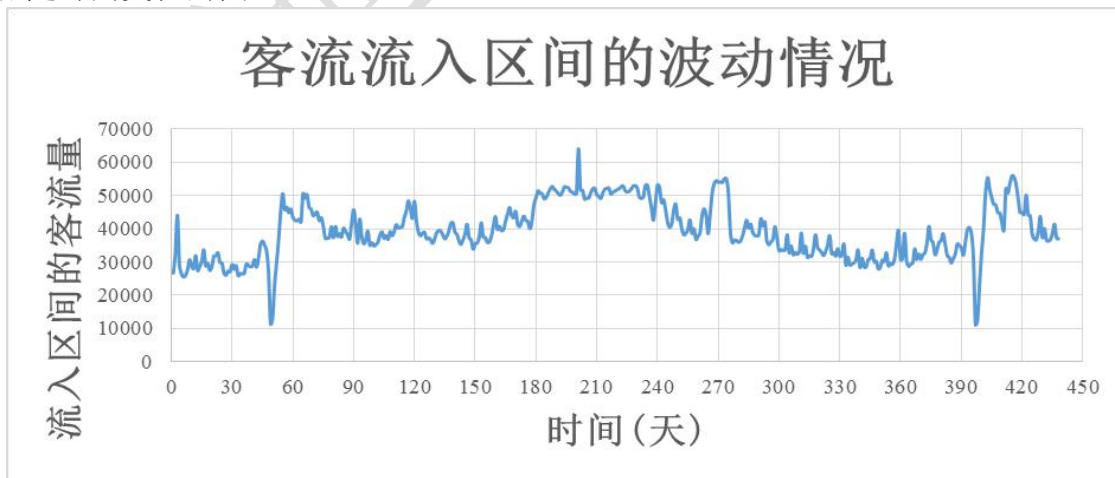


图 2: 流入区间客流随时间变化图

从图中明显看出，7,8 月份流入区间的客流较多，主要原因可能是 7,8 月份有较多的学生客流，且在节假日(如元旦，清明，五一，端午，中秋，国庆)都出现了客流的高峰

期，而两个春节的前夕客流量都出现了低谷，出现低谷的可能原因是春节之前旅游的人数相对较少。

②时段

针对一年的每天各个时段不同站点的上下车人数所做的统计，我们做出如下的图形：

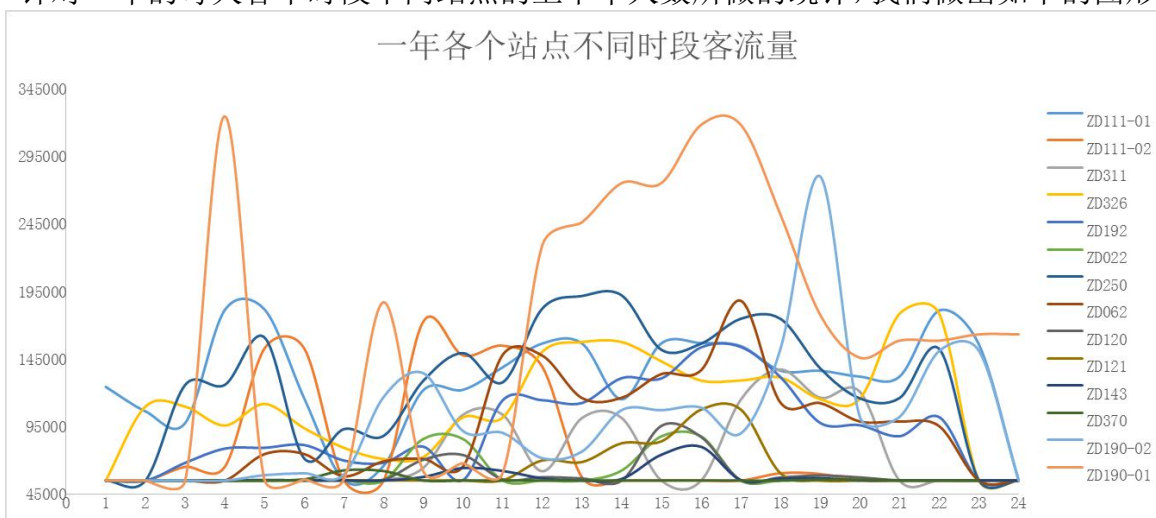


图 3：一年各个时段不同站点的上下车人数统计图

从上图中可清晰的看出各个站点一天中人流高峰期的时段，为此做出如下表格记录结果：

表 7：各站点一天人流高峰时段统计表

站点名	人流高峰期
ZD111-01	4:00-5:00 22:00-23:00
ZD111-02	5:00-6:00 9:00-10:00
ZD311	10:00-11:00 13:00-14:00 17:00-20:00
ZD326	12:00-14:00 21:00-22:00
ZD192	16:00-17:00
ZD022	9:00-10:00 15:00-16:00
ZD250	4:00-5:00 12:00-18:00
ZD062	11:00-12:00 16:00-17:00
ZD120	9:00-10:00 15:00-16:00
ZD121	16:00-17:00
ZD143	15:00-16:00
ZD370	7:00-8:00
ZD190-02	18:00-19:00
ZD190-01	4:00-5:00 7:00-8:00 16:00-17:00

根据图 3 和表 7 我们可直观的看出各个站点一天的客流分布规律，根据人流高峰时段的统计表，可看出一天大概有 4 个高峰时段：凌晨 4 点左右，上午 10 点左右，下午 4 点左右和晚上 10 点左右。该高峰期的分布具有一定的现实意义，通常路途较远时人们更倾向于早晨出发，能在天黑之前到达目的地，方便后面的活动安排。而短途区间内的乘车，人们会选择上午出发中午到达，处理完事物傍晚乘车，晚上回来。针对我们所统计的客流高峰时段，铁路部门可有针对性的对各个站点高峰时段进行客流疏导，增加站点的服务人员数，增强安保，维持车站的秩序。

③站点

根据表 4 的所统计的数据，我们可分析出一一年针对不同档次车各个站点的客流量高低，如下图所示。

运用改造的 RFM 模型，基于图 4 评估站点价值后将站点分区，分区原则如下图所示。A 类站点为重点保持站点，人流量大，档次高，只需维持现有规模即可，B 类站点虽然档次较高但人流量较小，因此需要重点建设站内设施，提高服务水平以期获得更高的人流量，C 类站点档次较低人流量较大，说明该站点存在潜在的升值空间，需要对该站进行扩建，提高站点的容纳量，以期更多的高铁动车经过该站，D 类站点档次较低人流量较小，说明该站点并没有很实际的存在价值，可考虑作为可取消站点。

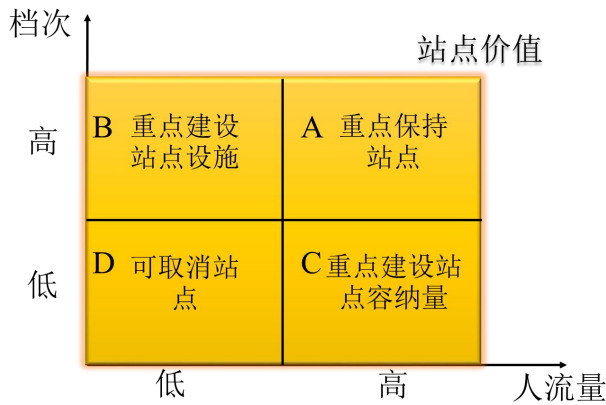


图 4: 针对各个站点分区原理图

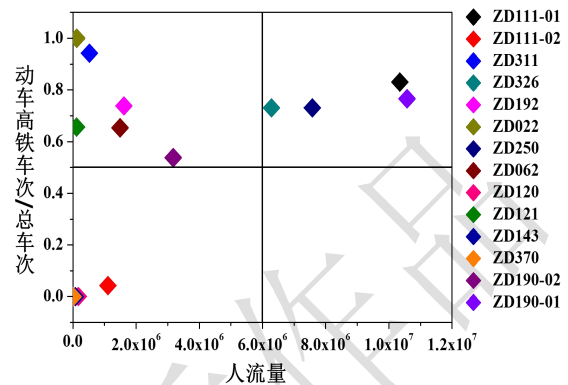


图 5: 各个站点分区图

基于上述模型和表 4，不难得到 14 个站点一年的人流量，档次高低通过经过动车高铁的车次与总车次的比值来衡量。从右上图中我们可以看出 ZD022 站点只有高铁动车，ZD120, ZD143, ZD370 站点只有普快特快和直达特快。这几个特殊的站点涉及铁路的运行规划问题，我们不能将其列入分区。ZD022 站点在 ZD250 市，是一个新规划的高铁站，以方便调度以及人口流动。ZD120, ZD143 和 ZD370 站点均在 ZD190 市，该市是一个大城市，这些站点很有可能是一些老旧的站点，只被规划为只停靠普通列车以及货车。此外，我们知道铁路上动车和高铁的优先级较高，这些站点很大可能作为普快的中途停车点，为动车和高铁让行。其他站点的分区结果如下表所示：

表 8: 站点的分区结果表

ZD111-01	ZD111-02	ZD311	ZD326	ZD192
A	D	B	A	B
ZD250	ZD062	ZD121	ZD190-02	ZD190-01
A	B	B	B	A

从上表可看出，A 类和 B 类站点较多，B 类站点表明这些站点很大可能是新增站点，站点人流量较小，但高铁和动车的车次较多。针对这些站点，我们可以重点进行站内设施建设，提高服务水平以吸纳更多的乘客。ZD111-02 可能是 ZD111 市的一个小站，通行的高铁动车不多且人流量不大，日后可作为可取消站点。

④区间

针对区间分析时，根据表 5 我们发现外地车从 ZD190-01 → ZD111-01 的车次较少，开车天数也较少，没有办法全面分析出客流流向规律，因此针对这种情况不做考虑，将区间分成三类：1)本地车从 ZD111-01 → ZD190-01; 2) 本地车从 ZD190-01 → ZD111-01; 3) 外地车从 ZD111-01 → ZD190-01。从区间的这三个角度分析客流规律。

1)本地车从 ZD111-01 → ZD190-01：我们通过观察表 5 的数据，做出各个区间内车上的总人数的变化图如左下所示：

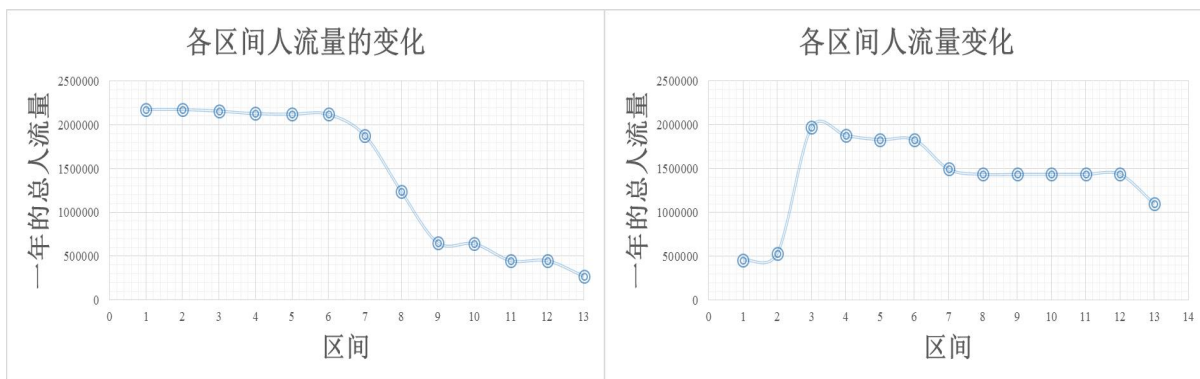


图 6: 本地车从 ZD111-01 → ZD190-01 和 ZD190-01 → ZD111-01 各区间人流量变化图

从上图中可看出，从 ZD110-01 发往 ZD190-01 的本地车，在前面两个城市 ZD110 市和 ZD326 市的客流量特别大，到 ZD250 市开始有较多人下车，其中在 ZD250 和 ZD062 站下车的人最多，ZD190 市的第一站 ZD120 站下车人数也特别多，之后客流量便较少，下车人流较多集中在 ZD190-02 站和 ZD190-01 站。客流流动规律的示意图如下所示：

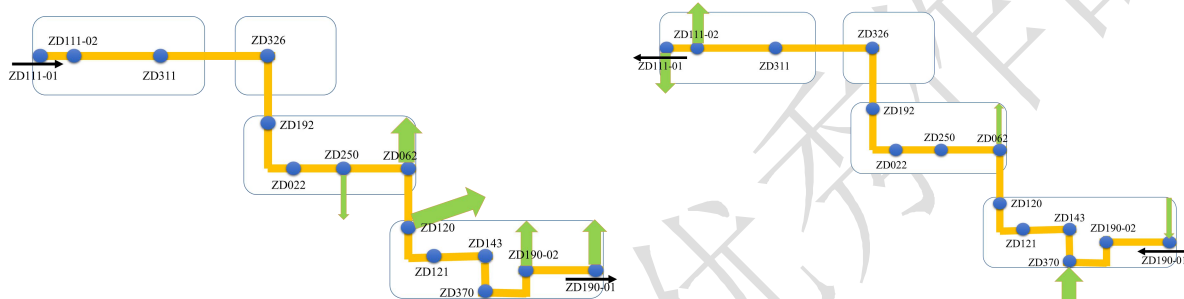


图 7: 本地车从 ZD111-01 → ZD190-01 和 ZD190-01 → ZD111-01 客流流动规律示意图

2) 本地车从 ZD190-01 → ZD111-01：同理我们也可做出各个区间内车上的总人数的变化图，从图中可看出：始发站 ZD190-01 的人不多，之后在 ZD370 站上车的人数特别多，之后主要的下车人数在 ZD250 市的 ZD062 站和终点城市的 ZD111-02 站和 ZD111-01 站。客流流动规律的示意图如图 所示。

3) 外地车从 ZD111-01 → ZD190-01：从区间车上人数的变化图可看出，乘客经过管内的初始站 ZD190-01 时客流量特别大，且在 ZD190-02 站仍有较大的上车人数，在 ZD326 市有部分人下车，在 ZD250 市下车人数特别多，主要集中在 ZD250 站和 ZD062 站，最后在 ZD190 市各个站点陆续有人下车。

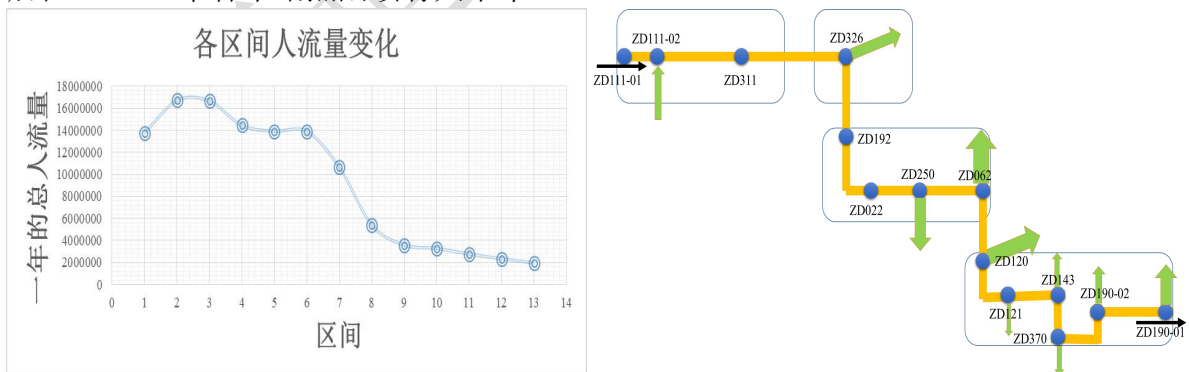


图 8: 外地车从 ZD111-01 → ZD190-01 各区间人流量变化图和客流流动规律示意图

总结:区间的客流规律如下：从 ZD111-01 → ZD190-01 的本地客流主要是去 ZD250 市和 ZD190 市。而外地客流较多的在 ZD250 市下车，ZD190 的各个站点都有下车人数。因此推测 ZD250 市和 ZD190 市是两个较大的旅游城市，经济发达。从 ZD190-01 → ZD111-01 的本地客流，更多的在 ZD190 市的 ZD370 站点上车，到终点站

ZD111 市下车。

步骤三：建模：

2.2.1. 基于时间序列的客流量预测模型

针对前面的客流规律的分析，建立基于时间序列的客流量预测模型，对客流量的数据进行分析，预测的结果以旅客列车梯形密度表的形式展示。下面以管外车到 ZD111-01 站点下车人数为例对其进行为期两周的预测，基于多元季节性时间序列(SARIMAX)的客流量预测模型建立的步骤如下：

①数据的预处理

通过题目所给的旅客列车梯形密度表，我们可以整理出如下表格的形式：

表 9：旅客列车梯形密度表

下车站	上车站	管外	ZD111	ZD311	ZD326	ZD192	ZD022	ZD250	ZD062	ZD120	ZD121	ZD143	ZD370	ZD190
管外														
ZD111														
ZD311														
ZD326														
ZD192														
ZD022														
ZD250														
ZD062														
ZD120														
ZD121														
ZD143														
ZD370														
ZD190														

首先针对管外车到 ZD111-01 站点下车人数一年多的数据进行整理作图如下：

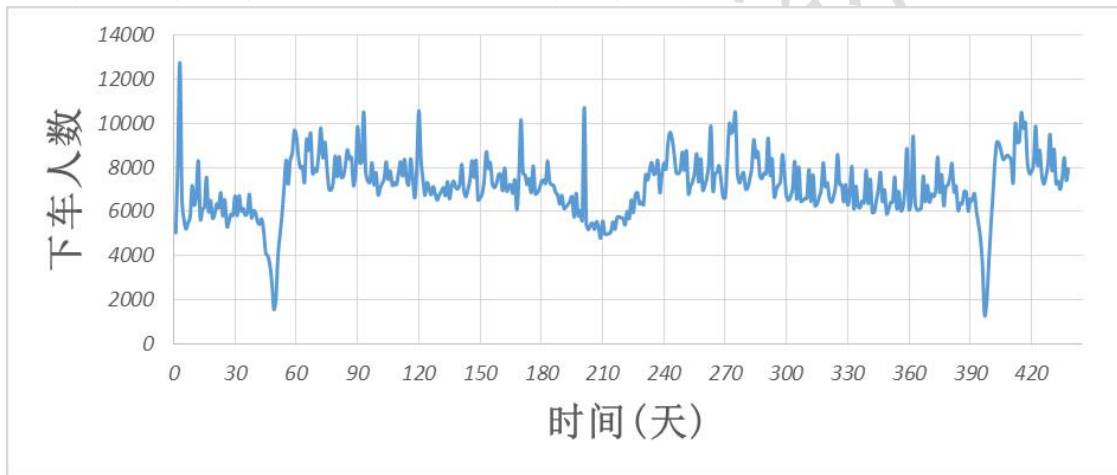


图 9：管外车到 ZD111-01 站点下车人数序列

进行小波分析后，将下车人数去毛刺，得到如下图的对照图：

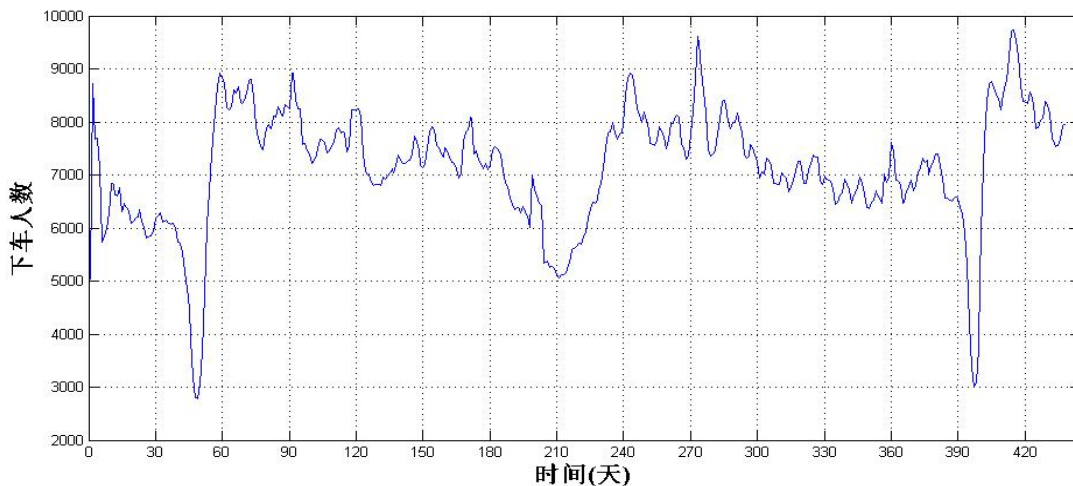
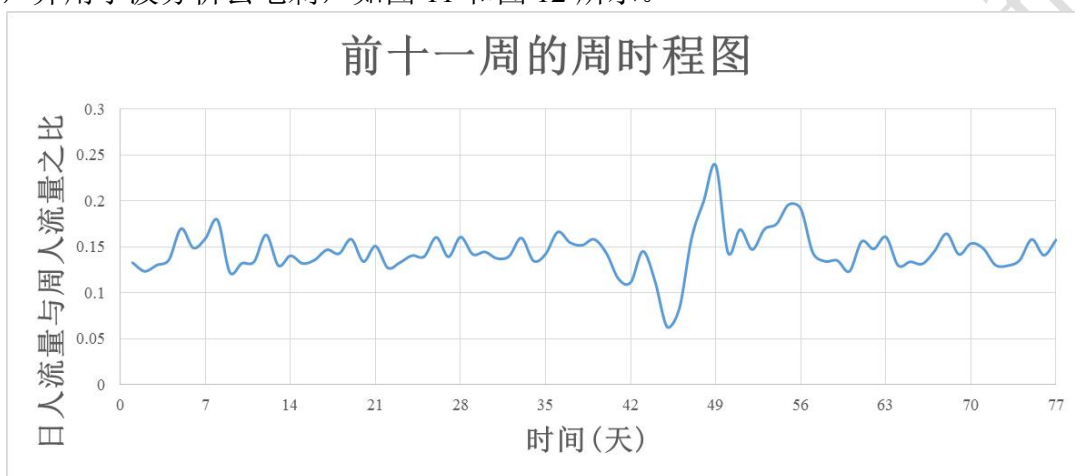


图 10: 小波分析后管外车到 ZD111-01 站点下车人数序列

上图显示的客流数据整体数据较为平稳，没有一个明显上升和下降的趋势，表明一年以来该站点客流量比较稳定。在图中可看出，1月初的时候存在一个人流的高峰期，此时正处于元旦的节假日，前面我们已经分析出 ZD111 市是一个旅游大市，自然会迎来较多的外地游客，这种人流高峰的情况也出现在清明和端午前后。而在两年的除夕前后都出现人流低谷期，可能是因为除夕人们都回家欢度春节，出来游玩的人相对较少。4 月份春运过后出现游客稀少的情况，而 7,8 月份客流又上升，这是由于暑假来临，高校学生离校或外来的务工人员转移引起的。

客流进行每周的分析,由于人们出行受到节假日的影响,客流在每周呈现出一定的变化趋势,其中周六、周天两天的客流量会有突然增加的趋势,国家法定节假日也会出现较大的客流波动。图中有一段时间的人流量波动特别大是由于春节。对于客流周的变化情况,本文引入日客流量与周客流量之比的客流周时程图。我们做出前十一周的周客流的周时程图，并用小波分析去毛刺，如图 11 和图 12 所示。



**图 11: 前十一周的周时程图
小波分析后的前十一周的周时程图**

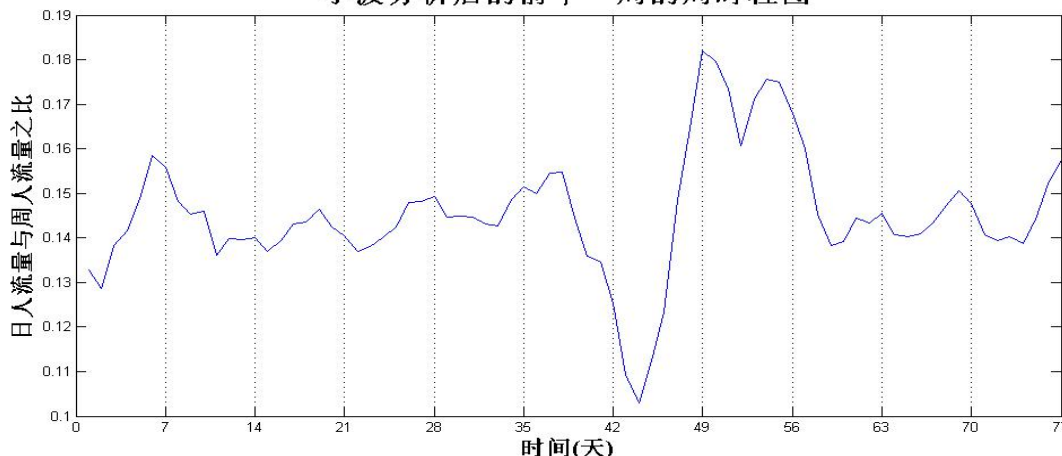


图 12: 小波分析后前十一周的周时程图

②模型的建立与求解

针对客流量预测的方法，一般基于 ARMA 的时间序列模型，其是在分析时间序列之间的相关性基础上，通过估计相关参数，采用历史数据预测未来发展趋势。相比于常规的时间序列方法，我们采用的多元季节性时间序列模型在考虑内生关系的基础上，分析外界变量对预测内容的冲击作用，因此大幅提高了外界条件变化时的预测精度”。相比于其他估计方法而言，多元季节性时间序列模型考虑了客流量随时间变化的波动性，且可量化评估外界因素对客流量波动造成的影响，易于工程应用。对于季节性多元时

间序列 SARIMAX(p, d, q) \times (P, D, Q) $_s$ 而言，常规的模型形态如下所示：

$$\phi(B)\Phi(B^s)\Delta_d\Delta_D^s(Y_t - \beta_0 - \sum_{k=1}^n \beta_k X_{k,t}) = \theta(B)\Theta(B^s)e_t$$

式中 Y_t 为 t 时刻的因变量； $x_{k,t}$ 为 t 时刻第 k 个解释变量； $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ 为标定参数； S 为周期差分长度； $\phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$ 为非季节性 p 阶自回归算子； $\phi_1, \phi_2, \dots, \phi_p$ 为对应的算子参数； $\Phi(B) = (1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_p B^{ps})$ 为 P 阶季节性自回归算子； $\theta_1, \theta_2, \dots, \theta_q$ 为对应算子参数； $\Theta(B) = (1 - \Theta_1 B^s - \Theta_2 B^{2s} - \dots - \Theta_Q B^{Qs})$ 为 Q 阶季节性移动平均算子； $\Theta_1, \Theta_2, \dots, \Theta_Q$ 为季节移动算子参数； $\Delta_d = (1 - B)^d$ 为非季节差分算子； $\Delta_D^s = (1 - B^s)^D$ 为 D 阶差分算子； B^i 为滞后算子； $B^i(Y_t) = Y_{t-i}$ ； e_t 为数据白噪声。其中，差分的目的在于使数据从非平稳数据变为平稳数据。以满足时间序列无偏估计的要求。

(1)数据的平稳性检验

考虑到数据存在时间趋势，具有一定的周效应，为分析该趋势是否明显，因此需要对原人流量序列进行自相关检验和单位根检验，利用 Eviews6.0 软件检验结果如下所示：

表 10：原人流量序列的单位根检验

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-6.662088	0.0000
Test critical values:		
1% level	-3.445127	
5% level	-2.867949	
10% level	-2.570248	

从检验的结果来看，该序列平稳，表明该序列没有明显的自相关性和周效应，但这并不代表每个站点的上下车人数都没有季节效应，所以我们设计的 SARIMAX 模型有实际意义。

(2)预测

利用 Eviews6.0 软件对其进行建模，先识别模型的阶数，做出如下的自相关函数和偏相关函数：

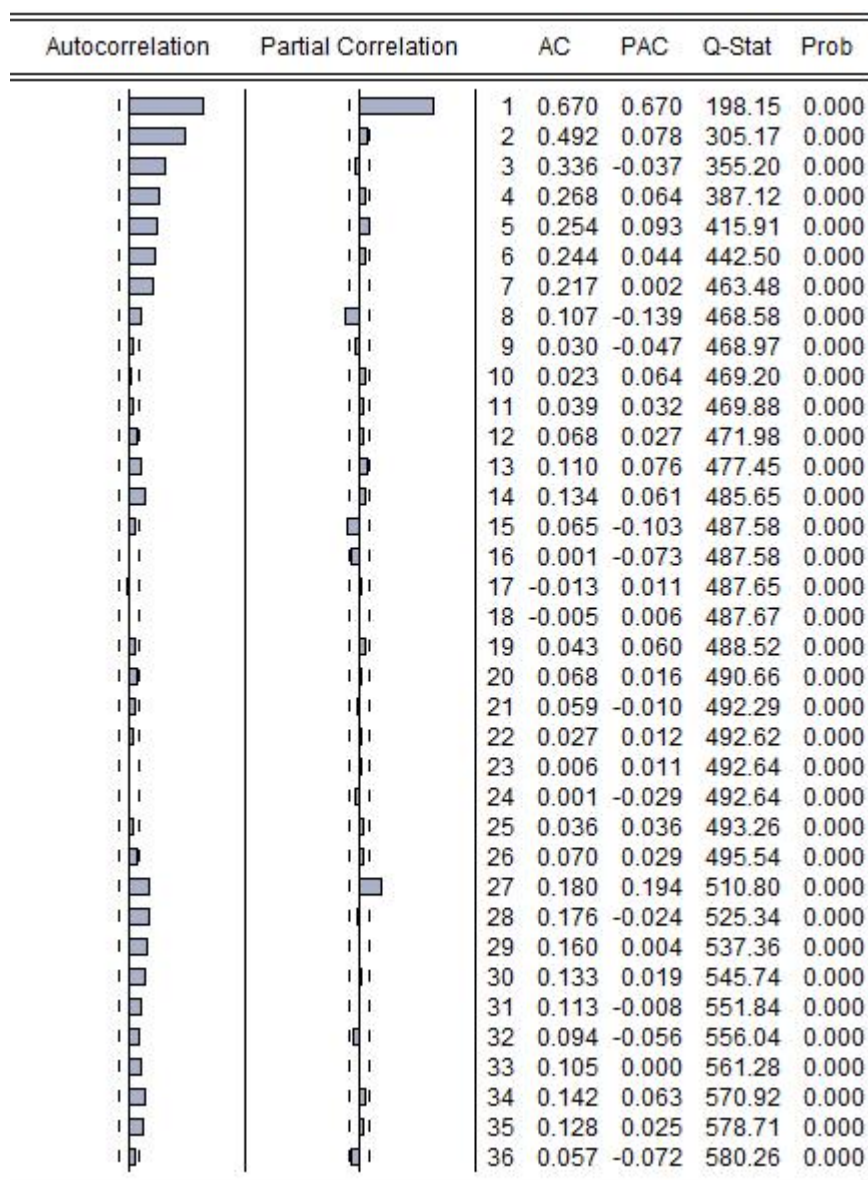


图 13: 自相关函数和偏相关函数图

从自相关函数和偏相关函数的滞后阶数可设计模型形式为 ARMA(4, 2), 运用该模型估计的结果为:

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1791.692	66.53897	26.92696	0.0000
AR(1)	1.337661	0.677129	1.975488	0.0489
AR(2)	-0.161037	1.048344	-0.153611	0.8780
AR(3)	-0.255725	0.345063	-0.741096	0.4590
AR(4)	0.056530	0.067825	0.833477	0.4050
MA(1)	-0.720201	0.678239	-1.061870	0.2889
MA(2)	-0.187415	0.641709	-0.292056	0.7704

考虑到预测的数据较多, 采用 MATLAB 编程统一求解, 至于 ARMA 自相关和偏相关的阶数问题, 循环遍历 1-10 阶, 找到最优的预测阶数。下面以管外 → ZD111-01 的下车人数预测为例, 作出原人数和预测人数的对比图如下所示:

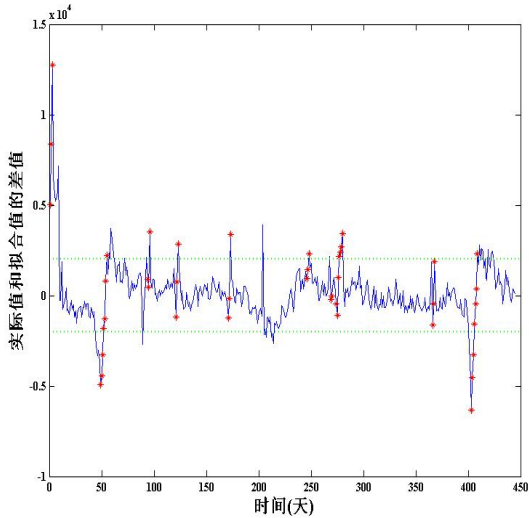


图 16: 实际值和拟合值的误差图

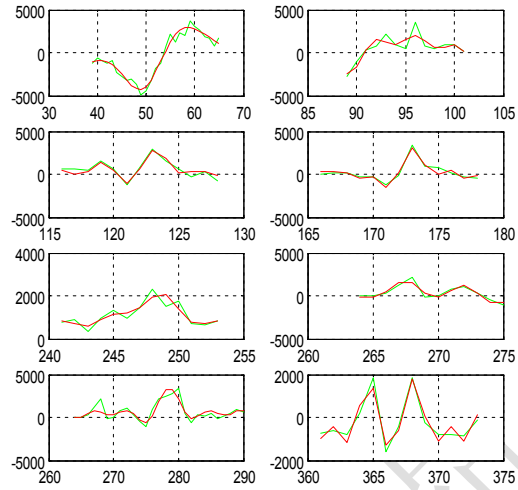


图 17: 各个节假日误差的傅里叶拟合

3) 数据分析发现 2: 预测值与实际值的线性关系

在数据分析过程中,我们发现用 SARIMAX 模型得出的每一天的预测值与误差值呈现简单的线性关系,尤其节假日,拟合优度相当高。

以清明节为例,以一年中同一天各个搭乘区间的人数的预测值为横坐标,实际值为纵坐标,得到下图:

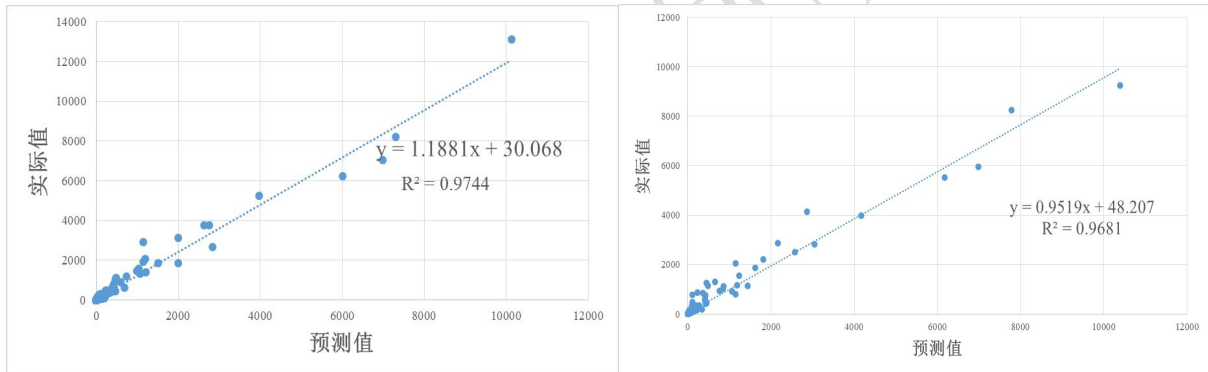


图 16: 清明节第一天和第二天的实际值与模型预测值之间的相关性图

可以看出预测值与实际值呈现很强的线性关系。

3)基于傅立叶拟合的SARIMAX模型节日误差修正技术

假设 1: 对一年中同一天不同的搭乘区间,实际客流量与预测客流量之间有线性关系,而不同搭乘区间的主要特点在于实际客流量不同,我们有理由认为我们所得到的线性关系表达式可以标注当天客流量与预测值的关系,而没有其他因素的影响(即得到的经验公式中,预测值是实际客流量的单变量函数)。

假设 2: 基于不同年份同一个节假日的放假天数一般相同,我们认为同一节假日的误差段仅仅由于客流量的不同而从振幅上进行改变,而频率不改变。

因此当我们考察不同年份的同一天时,对同一区间的客流量与预测值可以使用上述表达式。

对于两个不同的搭乘区间 $Interval1, Interval2$ 的客流量 p_1, p_2 , 按照我们的规律有:

$$\hat{p}_1 \approx kp_1 + b, \quad \hat{p}_2 \approx kp_2 + b$$

两式两边同时减去实际客流量

$$e_1 = \hat{p}_1 - p_1 \approx (k-1)p_1 + b$$

$$e_2 = \hat{p}_2 - p_2 \approx (k-1)p_2 + b$$

再将两式相减:

$$e_2 - e_1 = (k-1)(p_2 - p_1)$$

即

$$e_2 = (k-1)(p_2 - p_1) + e_1$$

节日误差段通常包括节日前中后几天的数据, 也即对于期间每一天 i 都可以得到一个对应的 k_i 。为了得到统一表征一个节日的 k , 我们将这些数据取均值:

$$k = \frac{1}{2d + dur(1..n)} (k_{dur(1)-d} + \dots + k_{dur(1)} + \dots + k_{dur(n)} + \dots + k_{dur(n)+d})$$

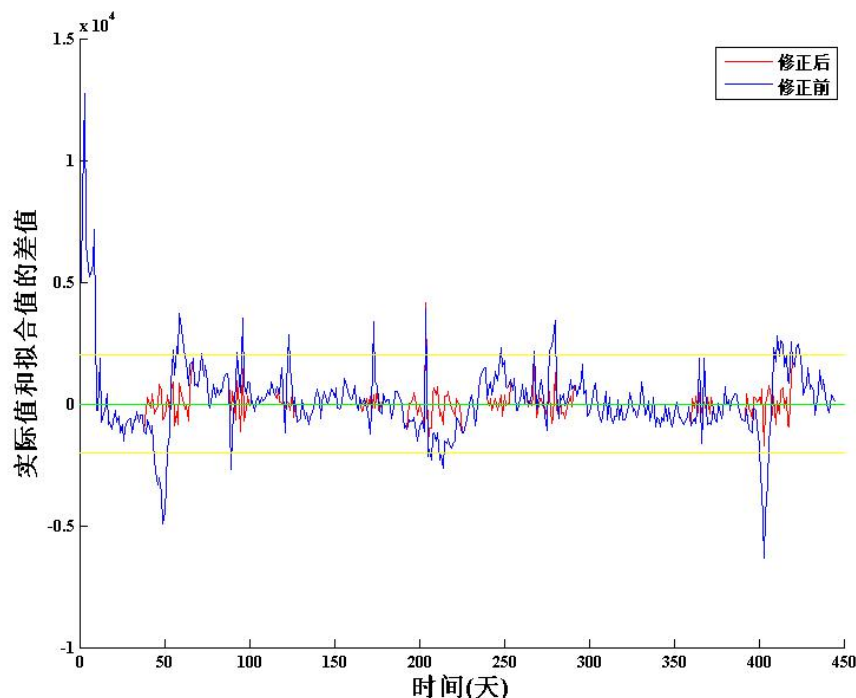
其中 d 表示节前节后的天数, 七天的假期取 $d = 10$, 三天的假期取 $d = 5$, $dur(1..n)$ 表示节日中的天数序号。

这样当我们对当前年份节假日误差进行修正时, 可以通过当前年份节前几天的客流量与已知年份客流量的差别估计出当前年份的节日误差段的变化情况, 也即得到当前年份的误差拟合函数。

4) 修正效果展示

统一使用三阶傅立叶拟合修正一年的误差, 得到各节日的误差修正拟合函数。再使用这种技术将前一年春节的误差函数修正后, 去拟合修正第二年春节, 效果较好。

基于上述误差修正原则, 我们得到最终的预测值和实际值的误差图如下所示:



从图中可看出，该修正方法十分合理，修正后的误差只在 0 点附近波动，基本消除节假日的影响因素。

(2)天气的影响因素

考虑天气的影响因素，选取两个变量，第一个变量是温度变量。由于温度通常有最高温度和最低温度两个表示方法，而题目中给的是一年的数据，最高气温和最低气温都会影响出行，因此我们采用最高温度和最低温度的均值来衡量。第二个表征天气特征的变量是天气状况。根据题目中所给的气象信息，每天的天气状况分别为晴天、晴到多云、多云转小雨、多云转阵雨、雷阵雨转阴、小雨、小雨到大雨，大雨到暴雨，霾，雨夹雪，雾，雨夹雪等级别。根据气象局对各种天气状况的分级，本文中天气分成 19 级，列表如下：

表 11: 气象局对各种天气状况的分级表

晴	多云	阴	阵雨	雷阵雨	雨夹雪	小雨
1	2	3	4	5	6	7
小到中雨	中雨	中到大雨	大雨	大到暴雨	暴雨	阵雪
8	9	10	11	12	13	14
小雪	中雪	大雪	雾	霾		
15	16	17	18	19		

针对第一个变量温度，作出拟合值与实际值之间误差随温度变化的误差图，从图中可看出其差值大多在 0 值附近波动，表明温度对客流量的影响不大。同理，我们也作出拟合值与实际值之间误差随天气分级变化的误差图，不难发现其差值大多在 0 值附近波动，表明天气对客流量的误差影响也不大。因此，在实际预测分析时，本文基于合理的分析忽略了天气的影响因素。

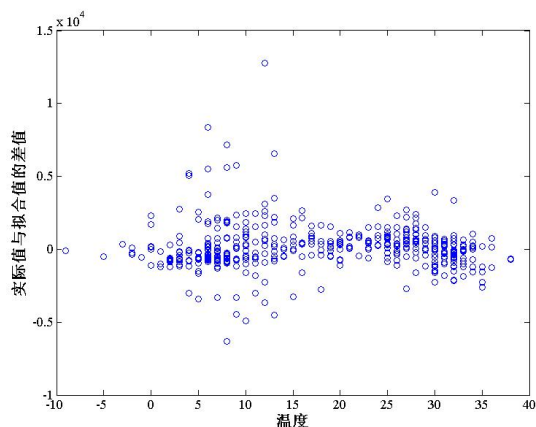


图 19: 实际值和拟合值的差值随温度变化图

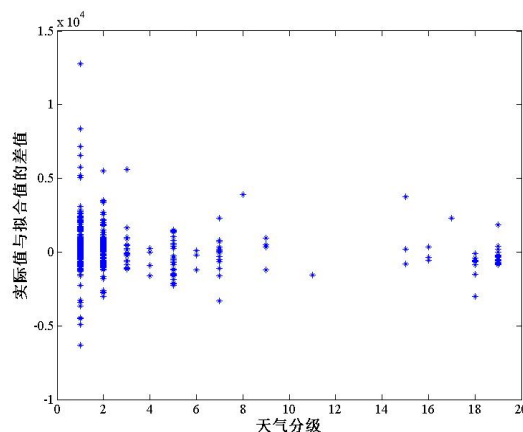


图 20: 实际值和拟合值的差值随天气分级变化图

步骤四：预测：

2.2.3.预测未来两周客流量

根据我们所建立的基于 SARIMAX 的客流量预测模型，并综合考虑了节假日的影响因素后，对 3 月 21 日-4 月 3 日的客流量进行预测。运用傅里叶拟合对 SARIMAX 模型在节假日的误差值进行修正，以 4 月 2 日为例，得到如下的预测结果：

	管外	ZD111-01	ZD111-02	ZD311	ZD326	ZD192	ZD022	ZD250	ZD062	ZD120	ZD121	ZD143	ZD370	ZD190-02	ZD190-01
管外	0	0	0	0	0	0	0	0	0	0	0	0	0	76	78
ZD110	10792	0	0	0	1212	392	0	1069	107	0	0	0	0	836	802
ZD311	246	991	0	0	0	0	0	0	0	0	0	0	0	0	31
ZD326	7245	3219	183	171	0	107	0	276	32	0	0	0	0	303	375
ZD192	1531	1025	37	91	373	0	0	180	0	0	0	0	0	44	136
ZD022	82	129	0	0	76	0	0	0	0	0	0	0	0	0	0
ZD250	8525	3240	213	157	1350	774	118	0	78	0	0	0	0	416	385
ZD062	515	425	55	48	189	65	0	1158	0	0	0	0	0	31	75
ZD120	179	0	95	0	63	0	0	131	0	0	0	34	0	99	66
ZD121	252	150	36	36	60	38	34	79	35	0	0	0	0	0	0
ZD143	53	32	49	0	44	0	0	49	0	37	0	0	0	74	59
ZD370	31	0	30	0	0	0	0	0	0	0	0	0	0	0	0
ZD190	15489	5776	336	310	2422	691	123	3651	1859	357	117	160	0	334	171

2.2.4. 优化车辆调度方案

① 基于多阶段决策的调度模型的建立

设第 k 天第 m 班次的车在各个站点上车人数构成的状态集合为 $X^{(k,m)} = [x_1^{(k,m)}, x_2^{(k,m)}, \dots, x_{14}^{(k,m)}]$, 第 k 天第 m 班次的车在各个站点下车人数构成的状态集合为 $Y^{(k,m)} = [y_1^{(k,m)}, y_2^{(k,m)}, \dots, y_{14}^{(k,m)}]$, 第 $i-1$ 个车站到第 i 个车站的里程为 a_i (其中 $i=1,2,\dots,14$)。

1) 目标函数

一年内列车总运营所获利润最高, 即一年中各个班次的旅客周转量最大,

$$\max P = \left\{ \sum_{k=1}^{445} \sum_{m=1}^{14} \sum_{i=1}^{14} (\sum_{s=1}^i (x_s^{(k,m)} - y_s^{(k,m)})) \cdot a_i \right\} \quad (1)$$

铁路局所购买的列车数 W 尽量的小, 即

$$\min W \quad (2)$$

2) 约束条件

① 上下行列车间的约束关系, 列车往返于两个站点之间, 设第 l 趟车发车时间为 t_l , 经过列车运行时间 Δt_1 , 在考虑短暂的休整时间 Δt_2 后的时刻将作为第 n 趟车发车时间 t_n 。

$$t_n = t_l + \Delta t_1 + \Delta t_2 \quad (3)$$

② 跨铁路局的长途列车作为约束不可改动的因素, 即第 l 趟车发车时间 t_l 与长途车经过该始发站的时间 t_j 间隔要大于最小行车间隔时间 Δt

$$|t_l - t_j| > \Delta t \quad (4)$$

注: 站点 ZD111-01, ZD111-02, ..., ZD190-01 依次记做站点 1, 2, ..., 14.

综上, 此规划模型可以写为:

$$\begin{aligned} \max P &= \left\{ \sum_{k=1}^{445} \sum_{m=1}^{14} \sum_{i=1}^{14} (x_i^{(k,m)} - y_i^{(k,m)}) \cdot a_i \right\} \\ \min W & \\ s.t. & \begin{cases} t_n = t_l + \Delta t_1 + \Delta t_2 \\ |t_l - t_j| > \Delta t \end{cases} \end{aligned}$$

② 基于客流量平均分配的 0-1 规划模型

根据模型中的约束, 我们发现求解该规划模型存在以下困难:

① 此规划模型是一个非线性规划, 很难得到全局最优解。

② 此规划数据量大, 要规划的内容包括列车停靠站点的设计, 发车时间的选取, 列车编组方案的调节, 难度较大。

因此, 我们对模型进行了改进和简化, 以日各车次的客座率平均分配为目标, 保证

日各车次总客座率达最大以设计各车次经过站点的分配。基于假设每个站点的客流量平均分配给每趟车，建立如下简化的规划模型：

设 $Lev(i, j)$ 为第 i 辆车从第 j 站离开时的人数； $Arv(i, j)$ 为第 i 辆车到达第 j 站时的人数； $X(i, j)$ 为第 i 辆车到第 j 站的停靠状态，0 表示不停，1 表示停； $Up(j)$ 为第 j 个站点上车的总人数； $Dn(j)$ ：第 j 个站点下车的总人数； $tn(j)$ ：第 j 个站点列车停靠数； $ut(j)$ ：第 j 个站点平均每辆车上车人数； $dt(j)$ ：第 j 个站点平均每辆车下车人数； $Avg(i)$ ：第 i 辆车的日均定员； $rate(i)$ ：第 i 辆车的客座率。

1) 目标函数

每天各个车次的客座率的方差最小，即各个车次的客座率尽量平均，

$$\min = \sum_{i=1}^9 \{rate(2i-1)\}^2 \quad (5)$$

2) 约束条件

① 基于假设每个站点日上车/下车总人数平均分配到该天经过该站点的所有车次的上车/下车人数

$$\begin{cases} ut(j) = Up(j) / tn(j) \\ dt(j) = Dn(j) / tn(j) \end{cases}, j = 1, 2, \dots, 14 \quad (6)$$

② 客座率的计算公式

$$rate(2i-1) = \sum_{j=1}^{13} Lev(2i-1, j) dis(j) / [Avg(i) dis(14)] \quad (7)$$

③ 第 i 趟车在第 j 个站点下车的人数要不小于该站点日平均每趟车下车人数

$$Arv(i, j) \geq dt(j), j = 1 \dots 14 \quad (8)$$

④ 第 i 趟车从第 j 个站点离开时的人数等于到达第 $j+1$ 个站点的人数

$$Arv(i, j+1) = Lev(i, j), j = 1 \dots 14 \quad (9)$$

⑤ 第 i 趟车在第 j 个站点离开时的人数等于该站点到达时的人数加减该站点上下车的人数

$$Arv(i, j) - dt(j) + ut(j) = Lev(i, j), j = 1 \dots 14 \quad (10)$$

⑥ j 站点与 $j+1$ 站点的距离为 $j+1$ 站点到起点与 j 站点到起点里程之差

$$dis(j) = Rt(j+1) - Rt(j), j = 1, \dots, 13 \quad (11)$$

⑦ 第 i 趟车在第 j 个站点经过上下车之后离开该站点时的人数小于该车次的日均定员

$$Lev(i, j) \leq Avg(i), j = 1 \dots 14 \quad (12)$$

⑧ 所有车次到第 j 站的停靠状态之和为该站点的日列车停靠数，注意这里不考虑第一站下车的情况

$$\sum_{i=1}^9 X(2i-1, j) = tn(j), j = 1, 2, \dots, 14 \quad (13)$$

注：站点 ZD111-01, ZD111-02, ..., ZD190-01 依次记做站点 1, 2, ..., 14.

综上，此规划模型可以写为：

$$\min = \sum_{i=1}^9 \{rate(2i-1)\}^2$$

$$\begin{cases}
 ut(j) = Up(j) / tn(j), j = 1 \dots 14 \\
 dt(j) = Dn(j) / tn(j), j = 1 \dots 14 \\
 rate(2i-1) = \sum_{j=1}^{13} Lev(2i-1, j)dis(j) / [Avg(i)dis(14)] \\
 Arv(i, j) \geq dt(j), j = 1 \dots 14 \\
 s.t. \begin{cases}
 Arv(i, j+1) = Lev(i, j), j = 1 \dots 14 \\
 Arv(i, j) - dt(j) + ut(j) = Lev(i, j), j = 1 \dots 14 \\
 dis(j) = Rt(j+1) - Rt(j), j = 1, \dots, 13 \\
 Lev(i, j) \leq Avg(i), j = 1 \dots 14 \\
 \sum_{i=1}^9 X(2i-1, j) = tn(j), j = 1, 2, \dots, 14
 \end{cases}
 \end{cases}$$

③基于 0-1 规划的站点分配设计

基于该 0-1 规划模型，运用 Lingo11.0 采用分支定界法求局部最优解，考虑到一年中节假日，周六周末的客流量较大，而工作日的客流量相对较小，因此，本文对站点优化调整时首先对不同时间段都进行讨论，最后综合考虑确定优化的方案：1) 工作日 2) 周六周日休息日 3) 节假日(包括元旦，春节，清明，五一，端午，胜利日，国庆)，整理结果如下：

1) 工作日

根据 Lingo 求得的局部最优解，我们可得到从 ZD111-01 → ZD190-01 的列车在工作日一天 9 趟车在各个站点的停靠车次，如下表所示：

表：ZD111-01 → ZD190-01 工作日各个站点的停靠车次表

ZD111-01	ZD111-02	ZD311	ZD326	ZD192	ZD022	ZD250
9	0	1	3	1	7	5
ZD062	ZD120	ZD121	ZD143	ZD370	ZD190-02	ZD190-01
3	0	1	0	0	7	9

2) 周六周日休息日

同理，可得到从 ZD111-01 → ZD190-01 的列车在周六周日休息日一天 9 趟车在各个站点的停靠车次，如下表所示：

表：ZD111-01 → ZD190-01 休息日各个站点的停靠车次表

ZD111-01	ZD111-02	ZD311	ZD326	ZD192	ZD022	ZD250
9	0	1	1	6	1	7
ZD062	ZD120	ZD121	ZD143	ZD370	ZD190-02	ZD190-01
9	0	9	0	0	9	9

3) 节假日

同理，可得到从 ZD111-01 → ZD190-01 的列车在周节假日一天 9 趟车在各个站点的停靠车次，如下表所示：

表：ZD111-01 → ZD190-01 节假日各个站点的停靠车次表

ZD111-01	ZD111-02	ZD311	ZD326	ZD192	ZD022	ZD250
9	0	1	3	6	6	9
ZD062	ZD120	ZD121	ZD143	ZD370	ZD190-02	ZD190-01
9	0	7	0	0	9	9

从 ZD190-01 → ZD111-01 的列车也类似分析，综合考虑工作日和节假日车次分配的问题，以及现有的各车次站点的停靠方案，我们做出部分的调整：目前 ZD250 站每个

车次均停靠，而根据我们的分析可知，ZD250 站的停靠车次可适当减少，综合考虑各车次该站点的客流量情况和发车时间等问题，最后决定将 D11 和 D18 班次的 ZD250 站点取消。增加 ZD062 站点和 ZD190-02 站点的设置，考虑各车次的客座率情况，增开 ZD062 站点的班次有 D16，增开 ZD190-02 站点的班次有 D05, D06, D07 和 D16。

③基于贪婪算法的编组方案设计

前面我们设计了站点的分配方案，而一天中可能出现的人流量很大，但某一趟车而言客座率却较低的情况，这时需要调整该车次的编组方案，具体步骤如下：

Step1: 初始化天数， $k = 1$ 。

Step2: 根据第 m 趟车在第 i 个站点的上下车人数状态 $X^{(k,m)}$ 和 $Y^{(k,m)}$ ，记录整天所有车次总客流量和各车次的客座率。

Step3: 判断 k 值：若 $k \neq 445$ ，则 $k=k+1$ ，并跳转到 Step2；若 $k = 445$ ，跳转到 Step4。

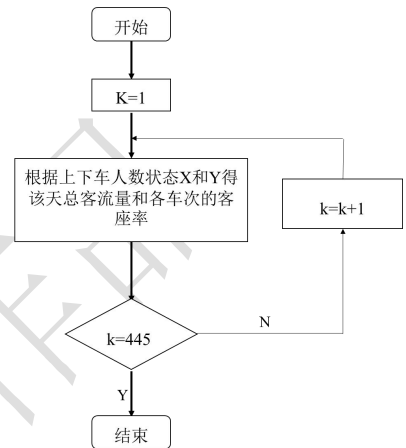
Step4: 结束循环，输出所有客流量和客座率信息进行分析。

具体流程图如右所示。

根据上述步骤流程图，我们采用人工搜索法，结合前面分析车次的客流规律，将 D05 车次由原先的 16 组编组改成 8 组编组。

④优化方案

基于上述的分析结果，我们给出最终的优化方案：取消 D11 和 D18 班次的 ZD250 站点；D16 增开 ZD062 站点；D05, D06, D07 和 D16 增开 ZD190-02 站点；D05 车次由 16 组编组改成 8 组编组。



2.3. 结果分析

将上述四个步骤得到的结果进行以下几个方面的详细分析：

1)对原始旅客列车梯形密度表中的数据进行挖掘分类整理，从四个角度(车次，时段，站点，区间)分析客流规律。

a.车次：不同车次的客座率有高低之分，其中，高铁各个车次的客座率普遍比较高；而普快客座率普遍较低；动车客座率部分车次相对较低，部分车次较高，例如：动车 D04 一年的整体的客座率都非常高，动车 D14 和 D16 一年的客座率都相对较低。2015 年对车次和部分车次的停车站点做了调整，调整之后的客座率得到了相应的提高，表明几次调整是合理的。通过比较管内客座率和整体客座率，发现 G07, G08 整年的管内客座率普遍低于总客座率，表明这两趟车有调整的空间。针对长途车，7,8 月份和元旦，清明，五一，端午，中秋，国庆前后流入区间的客流较多，而两个春节的前夕流入区间的客流较少。

b.时段：一天有 4 个高峰时段：凌晨 4 点左右，上午 10 点左右，下午 4 点左右和晚上 10 点左右。针对我们所统计的客流高峰时段，铁路部门可有针对性的对各个站点高峰时段进行客流疏导，增加站点的服务人员数，增强安保，维持车站的秩序。

c.站点：运用改造的 RFM 模型，将各个站点分区，其中，ZD022 站为新规划的高铁站，方便调度以及人口流动。ZD120, ZD143, ZD370 站只有普快特快和直达特快，可作为普快的中途停车点，为动车和高铁让行。ZD111-02 分为 D 区，通行的高铁动车不多且人流量不大，日后可作为可取消站点。其他站点设置都比较合适。

d.区间：从 ZD111-01 → ZD190-01 的本地客流主要是去 ZD250 市和 ZD190 市。外地客流较多去 ZD250 市和 ZD190 市。表明 ZD250 市和 ZD190 市是两个较大的旅游城市，经济发达。从 ZD190-01 → ZD111-01 的本地客流，更多的在 ZD190 市的 ZD370 站点上车，到终点站 ZD111 市下车。

2)建立 SARIMAX 预测模型,并综合考虑节假日和天气等因素,分析出节假日对客流产生明显影响,通过分析数据,运用经验公式,对原预测模型的数据进行修正。而天气因素的影响不明显,故本文对其忽略。

3)在站点客流平均分配到各个列车的假设下建立 0-1 规划模型和贪婪算法,得到各车次停靠站点和编组方案的优化设计。具体如下:取消 D11 和 D18 班次的 ZD250 站点;D16 增开 ZD062 站点;D05, D06, D07 和 D16 增开 ZD190-02 站点;D05 车次由 16 组编组改成 8 组编组。

3. 结论

总结本次比赛,我们根据题目所给的旅客列车的信息,深入挖掘已有数据,从车次,时段,车站和区间等多维度分析了客流规律。建立了 SARIMAX 时间序列模型,较为准确的刻画了客流规律,并综合考虑节假日和天气的相关因素,运用经验公式对原模型进行修正,较为合理的预测了两周的客流量。运用 0-1 规划模型和贪婪算法,得到各车次停靠站点和编组方案的优化设计。此外,本文基于题目中的数据设计了较为完整的数据库,方便以后的数据统计和分析。

4. 参考文献

- [1]茆诗松, 概率论与数理统计(第二版) [M], 高等教育出版社, 2013, 12.
- [2]刁在筠, 刘桂真, 运筹学(第三版) [M], 高等教育出版社, 2013, 12.
- [3]姜启源, 数学模型, 北京:高等教育出版社, 2011 年;
- [4]张晓峒:《Eviews 使用指南与案例》 [M],机械工业出版社,2007-02.
- [5]王洪业.基于客流预测的铁路旅客列车票额智能分配方法 [J].中国铁道科学.2003-05.