**2016年同济大学数学建模竞赛**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 姓名 | 学号 | 年级 | 所在专业 | 联系方式 |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

**游乐园客流疏导方案**

**摘要**

由问题（1）中的对游乐园里游客的游览提醒和疏导，我们考虑从两个方面进行考虑。一个方面为对游客的游览提醒，即对于任意某个位置的游客指引其想要去目的地的最短路径。另一方面为游客的疏导，即尽量使游乐园中各个项目点的负载程度保持相同，将较拥挤的项目点的游客转移引导到较稀疏的项目点。

针对对游客的游览提醒，我们比较了几种求路径的算法，最后利用现阶段比较成熟dijkstra算法做出求最短路径的模型，可以为项目点A、B…、J的游客做出他们想要去的目的地的最优的指导。

针对对游客的疏导，我们考虑到现实情况，根据现实中的语言变量引入语言算子，定义了表示了项目点当前负载程度的负荷率以及表示整个游乐园人分布是否均匀的均衡度。为疏引分散游客，提高整个游乐场的运行效率提供指导。

由问题（2）中的对于酒店的预测分析，我们考虑到预订酒店的影响因素有季节,周末,法定假日,暑期的影响，利用时间序列预测方法中的ARIMA（p，d，q）差分自回归模型，建立起对下一年酒店预订的模型。预测之后进行拟合的误差分析，得出拟合误差较小，该模型基本可以使用的结论。

**关键词**：最短路径、dijkstra算法、语言变量、负荷率、均衡度 、时间序列预测、 ARIMA模型

**问题一：**

1. **背景**

Youth游乐园，作为本市拥有最多过山车项目的游乐园，倍受年轻人的喜爱。为了让游客有一个快乐的游玩体验，需要对人流有一个合理的规划。

游乐园承载的客流量大小直接关系到游乐园的发展。游乐园的客流量过度稀疏，就会出现资源的闲置和浪费，使游乐园的经济收益下降，景区的收益过低不足以维持景区的正常运行的时候，旅游区也不能长久的发展下去。客流量的超载虽然会使短期的经济收益得到增加，但是人流的集中会导致一定区域的拥挤，会让游客的排队时间过长，甚至发生拥挤踩踏事件，所以我们要用合理的线路规划来使各区域的人流数量达到均衡，以避免事故的发生，改善游客的游玩体验，甚至可以增加游乐设施的寿命。给予游客最短路径可以让游客有更多的时间去游玩，不至于浪费时间在绕路上面，这可以让游乐园的游客循环的更快，从而增大了游乐园的载客量。

**二、问题重述**

Youth游乐园即将盛大开园，如何根据客流情况，及时分流人群，为顾客提供游园线路引导，保障游客的游园体验显得尤为重要。共设A-J 共10个项目点，游客可沿着图中标出的线路往返下个游乐项目。在保障每位游客体验游乐设施的前提下，建立对每个游乐项目的等候游客进行游览提醒和疏导的模型，以达到游园体验最优。基于以上可知我们需要解决的问题有：

1. 对任意位置的游客，我们为其想要去的目的地进行最短路径的指引，便于游客能够最方便快捷的到达其想要去的地方。
2. 对于游乐园的工作人员，我们要建立一个能够实时监测游乐园动态平衡度的模型，并显示各个项目点负载程度，为疏导游客提供帮助。

**三、符号说明**参数说明

|  |  |
| --- | --- |
| s | 起始点 |
| d（i） | 向量d表示起始点s到i区的最短路程距离和 |
| Index2(i) | 存放始点到第i点最短通路中第i顶点前一顶点的序号 |
| Pb(i) | 顶点i未标号为0，顶点已标号 为1 |
| Lu(n\*n) | 第i行表示从起始点s到i区的路径 |
| Ci | 各项目点更新游客的能力 |
| Li | 语言变量 |
| ρ | 负荷率 |
| σ | 均衡度 |
| λ | 游乐园的项目点均衡度阈值 |

**四：模型假设**

1. 只考虑项目点内的游客，对在项目间之间道路上的游客不予考虑。
2. 假设游乐园内各条道路相同，即只考虑道路的长度作为变量。
3. 假设游乐园中个项目点不会出现全部爆满的情况。
4. 假设游乐园中的各项目点的游乐设施是中间不简单连续运行，而且各个项目点开始和结束的时间相同。

**五、模型的建立与求解**

**5.1 模型一  
5.1.1最短路径** 最短路径问题是进行网络分析的核心问题之一。与网络相关的问题，如花费最小路径、次短路径、最大容量路径、最优路径和各种路径分配问题均可纳入最短路径问题的范畴之中，通过最短路径算法的探讨可以对如中心选址、资源分配、最大流量、最小费用算法等提供解决思路上的最优。在旅游线路的研究中最短路径的含义很广，不仅包括简单地理意义上的两点之间距离最短，还包括旅游活动的花费时间最短、费用最少和旅游区线路利用率程度最高等标准；同时最短路径还是旅游资源分配、游览路线设计及分析等线路优化问题的基础。此外，在旅游线路方面，如何根据实际景区景点的空间分布情况设计出比较合理的旅游线路，是一个需要深入研究的问题。最短路径的选择在交通线路分析和道路规划中有着非常重要的现实意义。国内外许多的专家学者在最短路径的选择即最短路径算法方面进行了大量的研究。经典的图论法与日渐完善的计算机数据结构及算法的有效结合使得新的最短路径算法不断涌现，且各具特色。本文选用的是选取现阶段应用较多且较为成熟的Dijkstra算法对最短路径求解。 **5.1.2 dijkstra算法简介**

Dijkstra算法**[1]**是由E.W.Dijkstra于1959年提出，又叫迪杰斯特拉算法，它应用了贪心算法模式，是目前公认的最好的求解最短路径的方法。算法解决的是有向图中单个源点到其他顶点的最短路径问题，其主要特点是每次迭代时选择的下一个顶点是标记点之外距离源点最近的顶点。

问题描述：在无向图 G=(V,E) 中，假设每条边 E[i] 的长度为 w[i]，找到由顶点 V0 到其余各点的最短路径。（单源最短路径） **2.算法描述**

1)算法思想：设G=(V,E)是一个带权有向图，把图中顶点集合V分成两组，第一组为已求出最短路径的顶点集合（用S表示，初始时S中只有一个源点，以后每求得一条最短路径 , 就将加入到集合S中，直到全部顶点都加入到S中，算法就结束了），第二组为其余未确定最短路径的顶点集合（用U表示），按最短路径长度的递增次序依次把第二组的顶点加入S中。在加入的过程中，总保持从源点v到S中各顶点的最短路径长度不大于从源点v到U中任何顶点的最短路径长度。此外，每个顶点对应一个距离，S中的顶点的距离就是从v到此顶点的最短路径长度，U中的顶点的距离，是从v到此顶点只包括S中的顶点为中间顶点的当前最短路径长度。

2)算法步骤：  
a.初始时，S只包含源点，即S＝{v}，v的距离为0。U包含除v外的其他顶点，即:U={其余顶点}，若v与U中顶点u有边，则<u,v>正常有权值，若u不是v的出边邻接点，则<u,v>权值为∞。

b.从U中选取一个距离v最小的顶点k，把k，加入S中（该选定的距离就是v到k的最短路径长度）。

c.以k为新考虑的中间点，修改U中各顶点的距离；若从源点v到顶点u的距离（经过顶点k）比原来距离（不经过顶点k）短，则修改顶点u的距离值，修改后的距离值的顶点k的距离加上边上的权。

d.重复步骤b和c直到所有顶点都包含在S中。

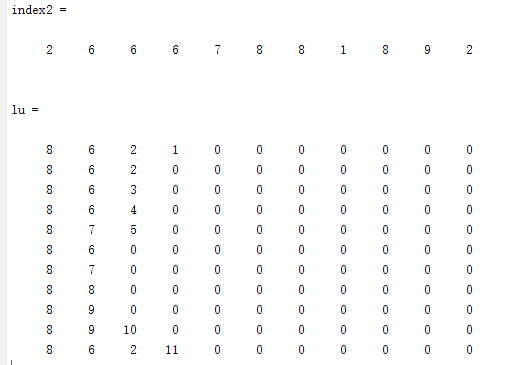
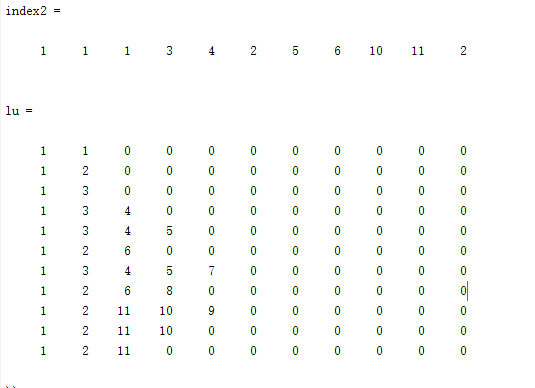
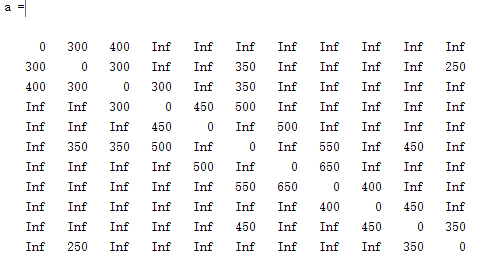
**5.1.3模型建立**

根据以上算法使用matlab 进行仿真，网络的连接用邻接矩阵an×n 表示,其中n 表示游乐园的项目数.a(i,j)表示 i项目到j项目的距离。当i和j无道路时，用Inf表示。当i=j时，a(i,j)=0,且a(i,j)=a(j,i)。该矩阵为11阶矩阵，包括出入口和10个游乐项目，1代表游乐园的出入口，2到11分别代表游乐项目A到J。比如a(5,8)表示5区（游乐项目D）到8区（游乐项目G）的距离。



|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 出入口 | 项目 A | 项目 B | 项目 C | 项目 D | 项目 E | 项目 F | 项目 G | 项目 H | 项目 I | 项目 J |
| 出入口 | 0 | 300 | 400 | Inf | Inf | Inf | Inf | Inf | Inf | Inf | Inf |
| 项目A | 300 | 0 | 300 | Inf | Inf | 350 | Inf | Inf | Inf | Inf | 250 |
| 项目B | 400 | 300 | 0 | 400 | Inf | 350 | Inf | Inf | Inf | Inf | Inf |
| 项目C | Inf | Inf | 300 | 0 | 450 | 500 | Inf | Inf | Inf | Inf | Inf |
| 项目D | Inf | Inf | Inf | 450 | 0 | Inf | 500 | Inf | Inf | Inf | Inf |
| 项目E | Inf | 350 | 350 | 500 | Inf | 0 | Inf | 550 | Inf | 450 | Inf |
| 项目F | Inf | Inf | Inf | Inf | 500 | Inf | 0 | 650 | Inf | Inf | Inf |
| 项目G | Inf | Inf | Inf | Inf | Inf | 550 | 650 | 0 | 400 | Inf | Inf |
| 项目H | Inf | Inf | Inf | Inf | Inf | Inf | Inf | 400 | 0 | 450 | Inf |
| 项目I | Inf | Inf | Inf | Inf | Inf | 450 | Inf | Inf | 450 | 0 | 350 |
| 项目J | Inf | 250 | Inf | Inf | Inf | Inf | Inf | Inf | Inf | 350 | 0 |

得到距离矩阵如下图  
在现实情况下，不仅要知道从出入口到各项目景点的最短路径，还有必要知道从任一项目景点到出入口以及其他项目景点的最短路径。我们设置参数s作为起始点，s可以取1到11.通过dijkstra算法求出s到其他区的最短路径。  
Matlab代码见附录。 **5.1.4求解模型** 使s=1,即从出入口出发，求到A到J区的最短路径。  
从图中可以看出从出入口到其他区的最短路径，比如到F区，则lu(7,:)为1,3，4,5,7.表示最短路径为出入口--A区—C区—D区—F区。  
 如果我们令s=8,即从G区出发，求到出入口或其他项目区的最短路径，其中1到8行分别依次表示从G区到A、B、C各个点以及出入口的最短路径。如下图所示：  
 **5.1.5模型分析：**



优点：从我们的模型可以看出，将起点设置为s,即可得到该点到所有点的最短路径，方便快捷的为每个位置的游客做出最优的线路，为游客提供便利。

缺点：虽然我们能求出任意位置到其他各个位置的最短路径，即单源最短路。但是若一些游客想制订一条可以经过多个项目点的最短路径，比如从A点到C点到H点再到J点的最短路径，利用我们这个模型就不能一次的快速求解，这也是Dijkstra算法的局限性。  
 **5.2 模型二  
5.2.1语言变量介绍：**

人们在日常生活中交流信息时使用的是自然语言，而自然语言的语义充满模糊性，为了对模糊的自然语言形式化和定量化，进一步区分和刻画模糊值的程度，常常借用自然语言中的修饰词**[2-3]**，譬如：“比较”“稍微”“有点”“非常”等，由此引入语言变量的概念。设*L* 为有限离散集合，且：*L* ={*lj*| *j* =-*t*,-(*t* -1),⋯,0,⋯,*t* -1,*t*}(*t* ∈*Z*∗且*t* ≥1)则称*lj* 为语言变量。其中，*Z*\* 表示正整数集。设-*t* 和*t* 分别表示在实际情况下决策者所使用的语言变量的下限和上限。比如，通过语言变量对某景区游客游览系统中某个景点进行拥挤度评价，集合{非常拥挤，拥挤，较拥挤，一般，较松散，松散，非常松散}可表示为*L*={*l*-3,*l*-2,*l*-1,*l*0,*l*1,*l*2,*l*3} 。为了获得更完整的语言信息，以及对语言算子进行运算，将*L* 扩展到一个连续的集合*L*ˉ，*L*ˉ={*l*ˉ*α* =-(*t* +1)<*α* <*t* +1,*α* ∈*R*}(*t* ∈*Z*∗且*t* ≥1), *L* ⊂*L*ˉ.

**5.2.2模型建立**

每个游乐项目的时间安排：



我们可以把游乐项目的每场可容纳游客数与每场持续时间之比作为游乐项目的更新游客的能力。经过变换之后以E为基准,以E的更新游客能力为1.000，得A、B、C、D、E、F、G、H、I、J更新游客能力用 Ci来表示，i=A、B…J

如下表：

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| CA | CB | CC | CD | CE | CF | CG | CH | CI | CJ |
| 0.606 | 1.200 | 1.000 | 0.600 | 1.000 | 1.000 | 0.75 | 1.000 | 0.667 | 1.250 |

在本文中项目点排队等候人数的语言变量我们通过游乐场管理人员在项目点的现场观察或通过各项目点布置的摄像头传回的图像数据从语言集中选取语言变量来描述。游乐区内有10个项目点，分别为A、B…、J。游乐区管理人员通过信息中心的监控平台同时对10个景点的状态进行监测，并通过语言术语集L={l-3 =‘非常松散’，l-2=‘松散’，l-1=‘较松散’，l0=‘一般’，l1=‘较拥挤’，l2=‘拥挤’，l3=‘非常拥挤’}中的自然语言对其进行描述。为了方便对语言术语集中的语言变量进行运算，将离散的语言术语集L-={-3，-2，-1，0，1，2，3}延拓到连续的语言术语集。

定义一：

项目点的负荷率：为游览景点排队等候的人数的语言变量与景点更新游客能力之比

即 ρ=Li/ Ci (1)

定义二：

游乐场的均衡度：为各景点负荷率的均方差。

即 σ=ρi-ρi）2（2）



由定义可知，项目点的负荷率表示了各个项目点的拥挤程度，项目点的负荷率越大表示越拥挤。

游乐场的均衡度表示了整个游乐的游客分布状态。当均衡度较大时，表示不同的项目点的排队人数差别很大，需要进行疏导。我们设立一个阈值λ，当均衡度大于λ时提醒工作人员当前游乐场的人数分配很不平均，需要进行疏导。

对游乐场游客的疏导可分为以下几步：

Step 1：通过信息中心摄像头采集的视频数据对游乐场内各景点的状态进行实时监控，采集项目点排队等

候人数的语言变量。

Step 2：通过式(2)计算出游乐区游客游览系统的均衡度并与阈值进行比较，一旦大于阈值λ，即进入Step 3。

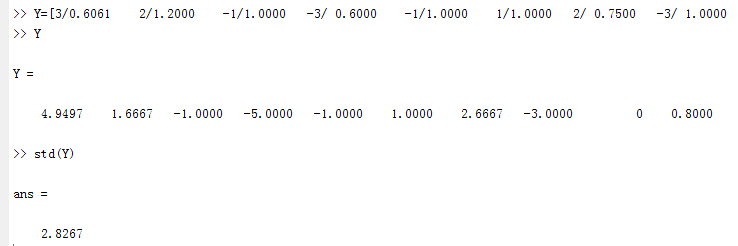
Step 3：通过式（1）算出各个项目点的负荷率，提醒高负荷率项目点的游客向低负荷率点进行转移。

**5.2.3计算实例：**

对于此游乐园，我们设立当10个点中最多出现‘拥挤’与‘松散’（即10个点之中不出现‘非常拥挤’与‘非常松散’）的均衡度的最大值为阈值，计算得阈值λ为2.5698。设某一时刻通过观测收集到A-I项目点的排队等候人数的语言变量为下表：

|  |  |
| --- | --- |
| A | 3 |
| B | 2 |
| C | -1 |
| D | -3 |
| E | -1 |
| F | 1 |
| G | 2 |
| H | -3 |
| I | 0 |
| J | 1 |

通过数值计算得景区的均衡度为：



均衡度为2.8267

均衡度大于阈值，表明对游乐场各项目点需要进行疏导。

再计算A 到 J 点的负荷率如下表：

|  |  |
| --- | --- |
| A | 4.9497 |
| B | 1.6667 |
| C | -1.0000 |
| D | -5.0000 |
| E | -1.0000 |
| F | 1.0000 |
| G | 2.6667 |
| H | -3.0000 |
| I | 0 |
| J | 0.8000 |

观察此表可得到A 、G 点人较拥挤，D 、H 点人稀疏，可告知游客各项目点的实际拥挤程度，并尽可能的使拥挤的点的人向稀疏点转移。

**5.2.4模型优缺点分析**

优点：引入了语言变量,贴近生活实际。定义了负荷率以及均衡度，比较有效的反映了游乐园各项目点的程度负载以及整个游乐园的人的分配程度，并且可以通过监控以及工作人员实时反映，便于管理。

缺点：只是粗略的使用负荷率的方差定义了均衡度，而且用游乐项目的每场可容纳游客数与每场持续时间之比作为游乐项目的更新游客的能力，虽然看起来比较合理，但是没有经过真实数据的检验，在实际情况中可能会有较大的误差。

**问题二：酒店预订预测  
一、背景：**

随着互联网以及手机APP的快速发展，人们利用互联网对生活提供便携帮助的例子越来越多，如提前订餐、订电影票、订酒店住房、预订出租车等等，为人们提供了极大的方便。对于酒店管理者来说，通过往年的预订情况来预测下阶段的预订情况，从而进行有效的管理和准备，比如增添设施及安保人员、提高酒店的管理，则显得尤为重要。本问题就是想解决通过过去一年酒店的预订住房情况，来预测下一年的预订情况，从而进行更好的酒店管理的例子。 **二、问题提出：**

皇冠假日酒店是游乐园内的酒店，目前已开业，为有需要的游客提供住宿便利。请根据该酒店历史预订数据信息,综合考虑影响房间预定量的主要因素(比如季节,工作日/周末,法定假日,暑期等)建立数学模型。并根据酒店2015年全年预定数据(附件2),预测2016年1月至3月每天预定房间数。

**三、问题分析：**

**3.1系统预测**

系统预测:根据系统发展变化的实际数据和历史资料，运用现代的科学理论和方法，以及各种经验判断，和知识，对，事物在未来一定时期内的可能变化情况，进行推测，估计和分析。

预测的特点

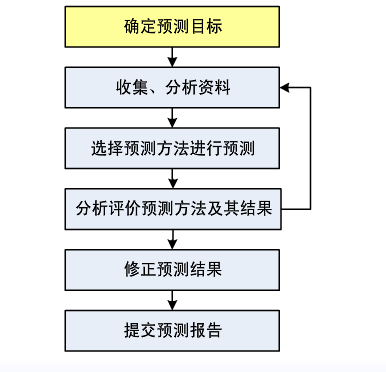
科学性:据统计资料和目前信息，运用一定程序方法和模型，分析预测对象与相关因素的互相联系，而解释预测对象特征和变形规律。

近似性:受许多随机因素的影响，提前预测的结果，往往与将来实际发生的结果有一定偏差。

局限性:对预测对象的认识常受知识，经验，观测和分析能力限制，又掌握资料和信息不够精确完整，或建模时简化等，导致预测的分析不够全面。

预测的顺序：

**3.2时间序列分析  
3.2.1 简介** 时间序列**[1.4]**:也叫时间数列、历史复数或动态数列。它是将某种统计指标的数值，按时间先后顺序排到所形成的数列。时间序列预测法就是通过编制和分析时间序列，根据时间序列所反映出来的发展过程、方向和趋势，进行类推或延伸，借以预测下一段时间或以后若干年内可能达到的水平。其内容包括：收集与整理某种社会现象的历史资料；对这些资料进行检查鉴别，排成数列；分析时间数列，从中寻找该社会现象随时间变化而变化的规律，得出一定的模式；以此模式去预测该社会现象将来的情况。  
**3.2.2 步骤**第一步



收集历史资料，加以整理，编成时间序列，并根据时间序列绘成统计图。时间序列分析通常是把各种可能发生作用的因素进行分类，传统的分类方法是按各种因素的特点或影响效果分为四大类：(1)长期趋势；(2)季节变动；(3)循环变动；(4)不规则变动。  
第二步

分析时间序列。时间序列中的每一时期的数值都是由许许多多不同的因素同时发生作用后的综合结果。

第三步

求时间序列的长期趋势(T)季节变动(s)和不规则变动(I)的值，并选定近似的数学模式来代表它们。对于数学模式中的诸未知参数，使用合适的技术方法求出其值。

第四步

利用时间序列资料求出长期趋势、季节变动和不规则变动的数学模型后，就可以利用它来预测未来的长期趋势值T和季节变动值s，在可能的情况下预测不规则变动值I。然后用以下模式计算出未来的时间序列的预测值Y：

加法模式T+S+I=Y  
乘法模式T×S×I=Y

如果不规则变动的预测值难以求得，就只求长期趋势和季节变动的预测值，以两者相乘之积或相加之和为时间序列的预测值。如果经济现象本身没有季节变动或不需预测分季分月的资料，则长期趋势的预测值就是时间序列的预测值，即T=Y。但要注意这个预测值只反映现象未来的发展趋势，即使很准确的趋势线在按时间顺序的观察方面所起的作用，本质上也只是一个平均数的作用，实际值将围绕着它上下波动。  
**3.2.3 分类** 时间序列预测法可用于短期、中期和长期预测。根据对资料分析方法的不同，又可分为：简单序时平均数法、加权序时平均数法、移动平均法、加权移动平均法、趋势预测法、指数平滑法、季节性趋势预测法、市场寿命周期预测法等。

简单序时平均数法也称算术平均法。即把若干历史时期的统计数值作为观察值，求出算术平均数作为下期预测值。这种方法基于下列假设：“过去这样，今后也将这样”，把近期和远期数据等同化和平均化，因此只能适用于事物变化不大的趋势预测。如果事物呈现某种上升或下降的趋势，就不宜采用此法。  
加权序时平均数法就是把各个时期的历史数据按近期和远期影响程度进行加权，求出平均值，作为下期预测值。

简单移动平均法就是相继移动计算若干时期的算术平均数作为下期预测值。

加权移动平均法即将简单移动平均数进行加权计算。在确定权数时，近期观察值的权数应该大些，远期观察值的权数应该小些。

上述几种方法虽然简便，能迅速求出预测值，但由于没有考虑整个社会经济发展的新动向和其他因素的影响，所以准确性较差。应根据新的情况，对预测结果作必要的修正。 **3.2.4 ARIMA预测** ARIMA模型全称为自回归积分滑动平均模型(Autoregressive Integrated Moving Average Model,简记ARIMA)，是由博克思(Box)和詹金斯(Jenkins)于70年代初提出一著名时间序列预测方法，所以又称为box-jenkins模型、博克思-詹金斯法。其中ARIMA（p，d，q）称为差分自回归移动平均模型，AR是自回归， p为自回归项； MA为移动平均，q为移动平均项数，d为时间序列成为平稳时所做的差分次数。所谓ARIMA模型，是指将非平稳时间序列转化为平稳时间序列，然后将因变量仅对它的滞后值以及随机误差项的现值和滞后值进行回归所建立的模型。ARIMA模型根据原序列是否平稳以及回归中所含部分的不同，包括移动平均过程（MA）、自回归过程（AR）、自回归移动平均过程（ARMA）以及ARIMA过程。  
基本思想：  
 ARIMA模型的基本思想是：将预测对象随时间推移而形成的数据序列视为一个随机序列，用一定的数学模型来近似描述这个序列。这个模型一旦被识别后就可以从时间序列的过去值及现在值来预测未来值。现代统计方法、计量经济模型在某种程度上已经能够帮助企业对未来进行预测。  
 如果假设随机变数Yt 为在时间t 的一个观测值, 那么一组Yt 所构成的数列就称为随机过程, 而一般标准的ARIMA (p , d , q)模型可以记为Yt ～ARIMA(p ,d ,q), 其定义为  
 φp(B)Wt = θq(B)At

式中:φp(B)= 1-φ1 B - … - φpBp ;Wt = (1-B)dYt ;θq(B)=1 -θ1B -θ2B2 -… -θqB q ;At 为白色噪音, 亦即At ～ N(0 , σ2a);p 、d 、q为非负整数;B 为后移分算子, 即BYt =Yt-1 ;φ1 , φ2 , …, φp 为自我回归参数, θ1 , θ2 , … , θp 为移动平均参数。

A RIMA 模型最大的优点在于对季节周期性数据指标的准确预测, 其与线性回归预测模型的区别为:线性回归预测模型的年度、季度、月度模型往往不能较好地揭示出被解释变量的非线性特征, 而A RIMA 模型的季度或月度模型能揭示出被解释变量的非线性特征;线性回归预测模型直接使用最小二乘法, 估计简单, 对含有解释变量的滞后项的回归模型, 则需要识别它的阶数, 而ARIMA 模型需要先估计它的阶数后, 再使用最小二乘法;利用线性回归预测模型进行预测时, 需要知道解释变量的预测值,而利用A RIMA 模型进行预测时不存在这个问题。  
ARIMA模型预测的基本程序

（一）根据时间序列的散点图、自相关函数和偏自相关函数图以ADF单位根检验其方差、趋势及其季节性变化规律，对序列的平稳性进行识别。一般来讲，经济运行的时间序列都不是平稳序列。

（二）对非平稳序列进行平稳化处理。如果数据序列是非平稳的，并存在一定的增长或下降趋势，则需要对数据进行差分处理，如果数据存在异方差，则需对数据进行技术处理，直到处理后的数据的自相关函数值和偏相关函数值无显著地异于零。

（三）根据时间序列模型的识别规则，建立相应的模型。若平稳序列的偏相关函数是截尾的，而自相关函数是拖尾的，可断定序列适合AR模型；若平稳序列的偏相关函数是拖尾的，而自相关函数是截尾的，则可断定序列适合MA模型；若平稳序列的偏相关函数和自相关函数均是拖尾的，则序列适合ARMA模型。

（四）进行参数估计，检验是否具有统计意义。

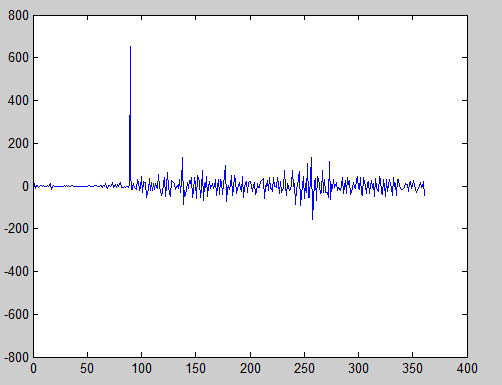
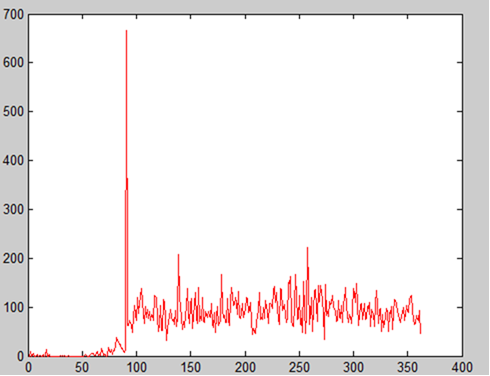
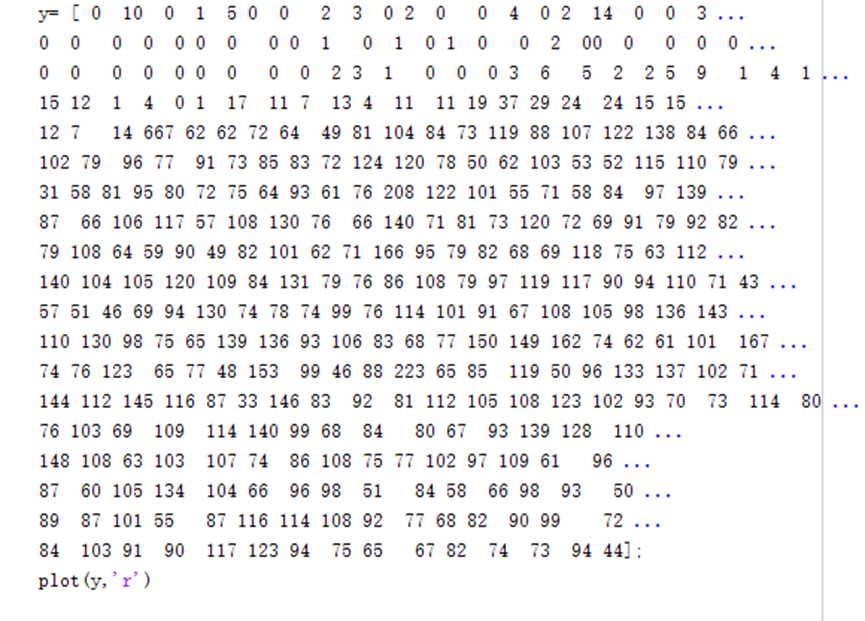
（五）进行假设检验，诊断残差序列是否为白噪声。

（六）利用已通过检验的模型进行预测分析。  
**四、 模型构架：  
4.1数据分析** 第1步用ARIMA预测要求序列是平稳的，若给出序列非平稳，则需对原数据进行预处理，使其平稳化．对数据求差分(一阶差分▽Zt=Zt-Zt-1，二阶差分▽2Zt=▽Zt-▽Zt-1)直到它是平稳的．此过程可以通过检查各种差分序列的相关图(包括自相关、偏相关图)直到找出一个“急速”下降于零，并且从此任何季节效应已经大大消除的序列来完成分析时间序列的随机性、平稳性及季节性．对于非季节数据，通常求一阶差分就足够了．对周期为12的季节数据，当季节效应是加性的，通常可以采用算子▽▽12；如果季节效应是乘性的，则可以采用算子▽212本身就足够了，不必外加差分．对于季节的数据，可以采用算子▽4等．

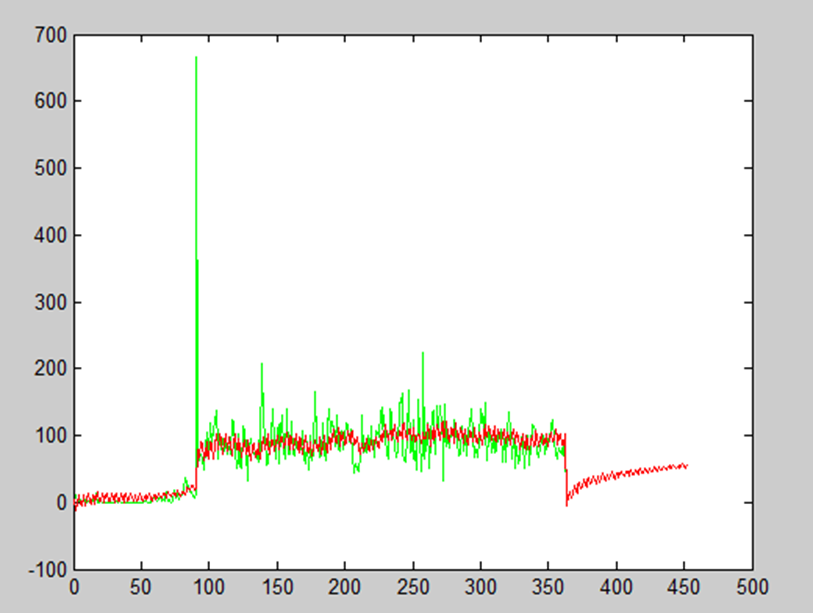
第2步选定一个特定的模型拟合所分析的时间序列．模型识别是建模环节中很重要的一环，是否合适的比较标准是：对一般ARMA模型体系中的一些特征，分析其理论特征，把这种选定模型的理论特征作为鉴别实际模型的标准，观测实际资料与理论特征的接近程度．最后根据这种分类比较分析的结果，来判定实际模型的类型．

第3步用时间序列的数据，估计模型的参数，并进行检验，以判定该模型是否恰当．如不恰当，则返回第2步，重新选定模型． **五、 模型求解：**

从图像可以看出，春季明显低于全年均值，其他三季相对来说比较稳定，因此考虑2016年春季的预订数时考虑季节影响。  
**5.1 序列平稳性的判断**  
根据酒店2015年全年的预订信息，画出散点图。  
  
  
  
图一 2015年酒店房间预订时间序列图  
  
  
  
  
图二 2015年酒店房间预订一阶差分时间序列图  
则一阶差分就很平缓，则d=1.  
**5.2模型识别及参数估计** 对于ARIMA模型，若自相关函数在滞后数为p后截尾和偏相关函数在滞后数为q后截尾，则阶数分别为p和q．经逐次尝试，得到各个ARIMA(p，1，q)模型的AIc参数，依据最小信息量原则选择ARIMA(3，1，3)做原序列预测的最优模型．AIC准则是1971年日本学者赤池(Akaike)给出的一种适用面非常广泛的统计模型选择准则，称为最小信息准则(Akaike Information Cri—terion)  
**5.3** **序列的预测** 通过分析2015年酒店房间预订数据，得到ARIMA的最优预测模型ARIMA(3，1，3)，该模型对原序列的预测效果图如图3(绿色为原始数据，红色为预测数据)所示，图3显示，预测值的趋势与实际值趋势基本保持一致．用ARIMA(3，1，3)模型预测2016年春季酒店预订，见表2，相对误差4.8％，在10％以内，可以认为模型预测结果准确，模型建立较好.  
  
运用算法原理编写matlab代码，代码见附录。



图三 模型原始数据与预测结果  
  
表二 2016年一月到三月预测：



|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 1st | 2nd | 3rd | 4th | 5th | 6th | 7th | 8th | 9th | 10th | 11th |
|  | 0 | 0 | 8 | 16 | 6 | 10 | 24 | 19 | 13 | 25 | 29 |
|  | 12th | 13th | 14th | 15th | 16th | 17th | 18th | 19th | 20th | 21st | 22nd |
|  | 19 | 23 | 34 | 29 | 23 | 31 | 37 | 26 | 30 | 40 | 33 |
|  | 23rd | 24th | 25th | 26th | 27th | 28th | 29th | 30th | 31st |  |  |
|  | 28 | 39 | 40 | 31 | 35 | 43 | 36 | 32 | 42 |  |  |
| 2 | 1st | 2nd | 3rd | 4th | 5th | 6th | 7th | 8th | 9th | 10th | 11th |
|  | 42 | 34 | 38 | 46 | 39 | 35 | 44 | 44 | 36 | 41 | 47 |
|  | 12th | 13th | 14th | 15th | 16th | 17th | 18th | 19th | 20th | 21st | 22nd |
|  | 41 | 38 | 47 | 46 | 39 | 43 | 49 | 42 | 41 | 48 | 47 |
|  | 23rd | 24th | 25th | 26th | 27th | 28th | 29th |  |  |  |  |
|  | 41 | 46 | 50 | 44 | 43 | 50 | 49 |  |  |  |  |
| 3 | 1st | 2nd | 3rd | 4th | 5th | 6th | 7th | 8th | 9th | 10th | 11th |
|  | 43 | 48 | 52 | 46 | 46 | 52 | 50 | 45 | 50 | 53 | 48 |
|  | 12th | 13th | 14th | 15th | 16th | 17th | 18th | 19th | 20th | 21st | 22nd |
|  | 48 | 54 | 52 | 48 | 52 | 55 | 50 | 50 | 56 | 53 | 50 |
|  | 23rd | 24th | 25th | 26th | 27th | 28th | 29th | 30th | 31st |  |  |
|  | 54 | 56 | 52 | 52 | 57 | 55 | 52 | 56 | 58 |  |  |

得到   
 d=1 p=3 q=3  
d为时间序列成为平稳时所做的差分次数  
p为自回归项  
q为移动平均项数  
ARIMA(3,1,3)模型。拟合误差为4.8%。  
**六、结果分析：** 从原始数据可以看出，酒店预订数受节假日，季节影响，在春季酒店预订数明显少于均值。进入夏季人数变多。同时预订数还受当地习惯等其他因素影响，在刚进入夏季甚至出现667人的全年高峰。模型考虑了春季低于均值，模型与原始数据也拟合的比较好，4.8%的误差也主要来自于数据偶然性。  
**七.模型优缺点：** 时间序列分析方法是一种很好的预测经济时间序列走势的方法。以上研究表明ARIMA（3,1,3）模型可以对酒店预订进行预测，预测效果较好，ARIMA可通过差分等数据转换方法将非平稳序列转变为平稳随机序列，以满足预测的前提，ARIMA使残差进入模型，提高了模型的精度，但是ARIMA建模法假定时间序列为未来的发展模式与过去的趋势一致，因此它往往只适用于作短期的预测．但总体来说，与其他方法相比，其预测的准确性还是比较高的，尤其是短期预测。 **附录  
1. 求两点之间最短路径Dijkstra算法matlab程序**

clear;

clc;

M=inf;

a(1,:)=[0,300,400,M,M,M,M,M,M,M,M];

a(2,:)=[zeros(1,2),300,M,M,350,M,M,M,M,250];

a(3,:)=[zeros(1,3),300,M,350,M,M,M,M,M];

a(4,:)=[zeros(1,4),450,500,M,M,M,M,M];

a(5,:)=[zeros(1,5),M,500,M,M,M,M];

a(6,:)=[zeros(1,6),M,550,M,450,M];

a(7,:)=[zeros(1,7),650,M,M,M];

a(8,:)=[zeros(1,8),400,M,M];

a(9,:)=[zeros(1,9),450,M];

a(10,:)=[zeros(1,10),350];

a(11,:)=zeros(1,11);

a=a+a'; s=1;

pb(1:length(a))=0;pb(s)=1;index1=s;

index2=ones(1,length(a));

d(1:length(a))=M;d(s)=0;temp=s;

while sum(pb)<length(a)

tb=find(pb==0);

d(tb)=min(d(tb),d(temp)+a(temp,tb));

tmpb=find(d(tb)==min(d(tb)));

temp=tb(tmpb(1));

pb(temp)=1;

index1=[index1,temp];

index=index1(find(d(index1)==d(temp)-a(temp,index1)));

if length(index)>=2

index=index(1);

end

index2(temp)=index;

end

lu=zeros(11);lu(s,[1 2])=[s s];

if s>1 && s<11

for i=[1:s-1 s+1:11]

tm=index2(i);

lu1=i;

while tm~=s

lu1=[tm lu1];

tm=index2(tm);

end

lu1=[s lu1];

lu(i,:)=lu(i,:)+[lu1 zeros(1,11-length(lu1))];

end

end

if s == 1

for i=2:11

tm=index2(i);

lu1=i;

while tm~=1

lu1=[tm lu1];

tm=index2(tm);

end

lu1=[1 lu1];

lu(i,:)=lu(i,:)+[lu1 zeros(1,11-length(lu1))];

end

end

if s == 11

for i=1:10

tm=index2(i);

lu1=i;

while tm~=11

lu1=[tm lu1];

tm=index2(tm);

end

lu1=[11 lu1];

lu(i,:)=lu(i,:)+[lu1 zeros(1,11-length(lu1))];

end

end

d

index1'

index2

lu

**2.ARIMA模型预测预订matlab代码**clear;

clc;

y= [ 0 10 0 1 5 0 0 2 3 0 2 0 0 4 0 2 14 0 0 3 ...

0 0 0 0 0 0 0 0 0 1 0 1 0 1 0 0 2 00 0 0 0 0 ...

0 0 0 0 0 0 0 0 0 2 3 1 0 0 0 3 6 5 2 2 5 9 1 4 1 ...

15 12 1 4 0 1 17 11 7 13 4 11 11 19 37 29 24 24 15 15 ...

12 7 14 667 62 62 72 64 49 81 104 84 73 119 88 107 122 138 84 66 ...

102 79 96 77 91 73 85 83 72 124 120 78 50 62 103 53 52 115 110 79 ...

31 58 81 95 80 72 75 64 93 61 76 208 122 101 55 71 58 84 97 139 ...

87 66 106 117 57 108 130 76 66 140 71 81 73 120 72 69 91 79 92 82 ...

79 108 64 59 90 49 82 101 62 71 166 95 79 82 68 69 118 75 63 112 ...

140 104 105 120 109 84 131 79 76 86 108 79 97 119 117 90 94 110 71 43 ...

57 51 46 69 94 130 74 78 74 99 76 114 101 91 67 108 105 98 136 143 ...

110 130 98 75 65 139 136 93 106 83 68 77 150 149 162 74 62 61 101 167 ...

74 76 123 65 77 48 153 99 46 88 223 65 85 119 50 96 133 137 102 71 ...

144 112 145 116 87 33 146 83 92 81 112 105 108 123 102 93 70 73 114 80 ...

76 103 69 109 114 140 99 68 84 80 67 93 139 128 110 ...

148 108 63 103 107 74 86 108 75 77 102 97 109 61 96 ...

87 60 105 134 104 66 96 98 51 84 58 66 98 93 50 ...

89 87 101 55 87 116 114 108 92 77 68 82 90 99 72 ...

84 103 91 90 117 123 94 75 65 67 82 74 73 94 44];

plot(y,'g')

hold on

and1=0;and2=0;

for k=1:90

and1=and1+y(k);

end

and1=and1\*1.0/90;

for k=91:362

and2=and2+y(k);

end

and2=and2\*1.0/272;

and3=and2-and1

y=y';

Data=y;

SourceData=Data(1:362,1);

step=91;

TempData=SourceData;

TempData=detrend(TempData);

TrendData=SourceData-TempData;

H=adftest(TempData);

difftime=0;

SaveDiffData=[];

while ~H

SaveDiffData=[SaveDiffData,TempData(1,1)];

TempData=diff(TempData);  
 difftime=difftime+1;  
 H=adftest(TempData);

end

u = iddata(TempData);

test = [];

for p = 1:5

for q = 1:5   
 m = armax(u,[p q]);

AIC = aic(m);   
test = [test;p q AIC];

end

end

for k = 1:size(test,1)

if test(k,3) == min(test(:,3))   
p\_test = test(k,1);

q\_test = test(k,2);

break;

end

end

TempData=[TempData;zeros(step,1)];

n=iddata(TempData);

m = armax(u,[p\_test q\_test]);

P1=predict(m,n,1);

PreR=P1.OutputData;

PreR=PreR';

if size(SaveDiffData,2)~=0

for index=size(SaveDiffData,2):-1:1

PreR=cumsum([SaveDiffData(index),PreR]);

end

end

mp1=polyfit([1:size(TrendData',2)],TrendData',1);

xt=[];

for j=1:step

xt=[xt,size(TrendData',2)+j];

end

TrendResult=polyval(mp1,xt);

PreData=TrendResult+PreR(size(SourceData',2)+1:size(PreR,2));

tempx=[TrendData',TrendResult]+PreR; tempx(1,363:453)=tempx(1,363:453)-and3;

plot(tempx,'r');

tempx(1,363:453)

参考文献

[1] 姜启源，数学模型，北京：高等教育出版社，2011

[2] 方荣华．基于ARMA时间序列理论的建模及应用阴．科技资讯，2012，(19)：197．198．

[3] Crampon L J．Gravitational model approach to travel market analysis[J]．Journal ofMarketing，1 966，30(2)：27～3 1．

[4] 王仁志, 苗维亚. 基于M/M/n 排队论的大型旅游景区内部排队现象研究[J]. 经济体制改革, 2012, (3): 177-180.

[5] 郭娟. 旅游景区拥挤问题理论分析与解决方案研究——以五台山景区为例[J]. 山西农业大学学报:社会科学版, 2010, (4): 476-479.