

# 航班延误问题

陈芷君<sup>1</sup> 龚芳逸<sup>1</sup> 王瑞<sup>1</sup>

指导教师 李工农<sup>1</sup>

1. 深圳大学 数学与计算科学学院 深圳市 518000

## 摘要

中国民航业近年来飞速发展,航班量增多、航班密度逐步加大,而机场大面积延误、各机场准点率低下、资源配置矛盾等问题也日益凸显出来,香港南华早报甚至公开报道称中国的航班延误最为严重。针对上述问题,本文致力于探索其真实性,究其本源并制定改善方法。首先,针对香港南华早报所报道的中国机场准点率最低的结论<sup>[1]</sup>展开讨论,探究结论的真实性。第一,在flightstats.com对中外机场延误统计方法存在双重标准的猜想假设下,通过控制变量的方式,将美国航空数据网站flightstats.com在2013年公布的数据<sup>[2]</sup>进行统一标准处理,以30分钟内为准点的标准进行了重整计算,将其所占百分比换算成新的准点率与国内的统计结果<sup>[3]</sup>进行了对比,验证结果表明统计过程中不存在双重标准。第二,由于上述猜想不成立,故根据上述对比结果再次提出数据估计过程中误差较大的猜想。经过数据验证,由于估算滑行时间小于实际滑行时间,确实造成了大量的延误统计量的出现,从而导致了我国的机场放行正点率一直处于垫底位置。故对于以上猜想验证,本文将原数据重新处理排列,证实报道结论与事实不符。在国内外统计方式不同的情况下,以直观的延误率来说明机场放行延误状况并不能使双方信服,因此运用熵权法和层次分析法,建立以影响延误率的因素为指标的新的评价体系,来对各机场的延误状况进行排名。

其次,对于航班延误的问题根源,由于天气问题可以直接影响航班的延误,故本文在假定天气为主要因素的条件下,找寻其他可控原因。通过对机场数据的筛选,运用 R 语言编程做最优子集回归,选取了货邮吞吐量、旅客吞吐量以及起降架次和平均出港滑行时间、日均起降架次、日高峰架次以及放行正常率和 GDP 为变量,筛选结果证明各自变量与因变量间具有强相关性,说明变量选择有效。然后为使主要原因独立,验证各变量间的相关性,运用SPSS做主成分分析,总结得到我国航班延误的主要原因为机场原因和空管原因。为验证总结原因合理化,通过纵向对比不同年份的变化,发现与分析结果基本相符。最后,为进一步证实结论的可靠性,与民航局发布的年报进行了对比,发现本文结论与其基本一致。

通过查阅资料,我国每年因航班延误造成了约 500 亿元的损失。本文以延误所致经济损失最小为目标函数,使用 MATLAB 利用蚁群算法建立模型,通过优化航班调度,来减少延误时间,缩小经济损失,改善航班延误问题。我们以深圳宝安机场为实例进行了计算,结果表明我们提出的模型能为深圳宝安机场每年节约 15 亿元人民币左右。

关键词: 层次分析法 熵权法 R语言 最优子集回归 主成分分析 MATLAB 蚁群算法

## 一、问题重述

机场延误一直是热点话题。香港南华早报网于2013年7月12日根据flightstats.com2013年6月份的统计称：中国的航班延误最严重，国际上航班延误最严重的10个机场中，中国占了7个。其中包括上海浦东、上海虹桥、北京国际、杭州萧山、广州白云、深圳宝安、成都双流等机场<sup>[1]</sup>。然而，对比《2013 年全国航班运行效率报告》<sup>[3]</sup>，这些机场的放行准点率均远高于 flightstats.com 所统计的。针对这一矛盾，为找出差别原因，求得较准确的结论，探寻我国航班主要延误原因，寻找改进方法，我们要解决的问题是：

1. flightstats.com 的统计是否正确？中国机场在国际上延误率是否真这么低？
2. 我国航班延误的主要原因是什么？
3. 对航班延误有什么改进措施？

## 二、问题分析

### 2.1 对问题一

#### (1) 对报道中结论的质疑与猜想分析

美国航空数据网站flightstats.com的2013年6月份的《Airline and Airport On-time Performance Report》中显示，从北京首都国际机场起飞的 22019 个航班中，仅有 18.30% 准时起飞<sup>[2]</sup>，2013年全年从北京国际机场起飞的航班中，仅有31.78%准时起飞，而根据中国民航统计局发布的《2013 年全国民航航班运行效率报告》，北京首都国际机场2013年的放行不正常率(不正常情况包括延迟起飞和取消航班)为33.67%<sup>[3]</sup>，即放行准点率约为 66.33%。

以上两者对同一机场准点率公布的统计数据相差如此之大，不得不让我们产生质疑，是什么导致了这样的差别呢？在假定数据真实的条件下，对于这种差别的产生我们有两种猜想，分别为在统计方式上存在双重标准以及数据处理上的问题这两个方面。

猜想一：双重标准

对比国内外判断延误标准我们可以将复杂的飞机起飞过长进行简化，为方便表述我们规定：

时间点 A：计划关舱时间（也就是机票上标注的时间、计划起飞时间）<sup>[4]</sup>

时间点 B：实际关舱时间

时间点 C：实际起飞时间

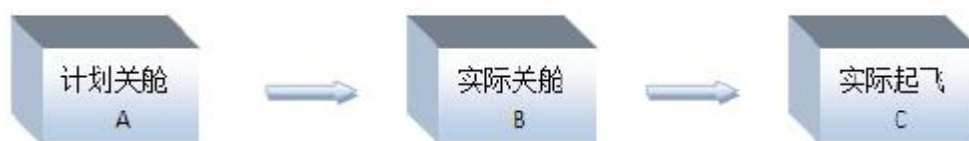


图 2.1 飞机起飞流程

对于延误统计的统计方法，国内外的标准是不同的，美国是以  $B-A > 15\text{min}$  即为延误统计的，而中国是以  $C-A > 30\text{min}$  为延误统计的。flightstats.com 在公布数据后引起了网上的热议，其中有一种被广泛网民认可的说法是 flightstats.com 在做统计排名时对中国和其他国家采取的是不同的统计标准，即同以超 15min 算延误，对其他国家按照  $B-A$  统计，而对我国则按照  $C-A$  统计。据此，我们考虑通过验证 flightstats.com 的统计是否存在双重标准来作为判断其报告排名是否正确的方法之一。由于 2013 年的每班航班的具体时间点 A、时间点 B、时间点 C 不可考，所以，不能从数据上直接证明有双重标准。我们假定

flightstats.com 网站从 2013 年到 2015 年并没有修正统计方法，拟选取 2015 年近期航班数据，分别就计划起飞时间和实际起飞时间在 flightstats.com 网站和国内应用非常广泛的商业网站“飞常准”进行对比，来验证是否存在双重标准问题。

为进一步验证（即假设 2013 年与 2015 年是采取的不同统计方法），我们拟将 2013 年全年的机场延误率数据进行重整，来推断 2013 年是否采取了双重标准。即对问题一：

若网站未采用双重标准（即对国内外的航班均按照 B-A 来统计，超过 15min 为延误），我们假设其 30min 以内起飞均为正点，由于 B 至 C 间还有飞机滑行时间，则 flightstats.com 的正点率应高于我国统计的正点率。

具体说明如下（假设国内外均以 30min 以内为正点，国外以 B-A 计算，国内以 C-A 计算，延误算统计量 0，否则算统计量 1）：

表 2.1：延误统计量统计方法

	国外	国内
$B-A < 30\text{min}$ 但 $C-A > 30\text{min}$	1	0
$B-A > 30\text{min}$	0	0
$C-A < 30\text{min}$	1	1

因此，若未采用双重标准，国外的“准点率”应高于国内的准点率。若采用了双重标准（即算国内的航班为 C-A），则和国内统计标准一样，那么国外的“准点率”应该约等于国内的准点率。

猜想二：数据处理出问题

我们猜测由于航空公司数据机密等原因，导致 flightstats.com 网站并未取得实际关舱时间数据，在为了统一标准的情况下对所得数据进行一定处理，导致数据失真。

（2）建立新的体系：

由于统计方式的不同，直接观测延误率不能令双方信服。因此构建不以延误量而以影响延误的因素为指标的新的评价体系。

根据民航局发布的年报中所公布数据以及第二问的结论，在谨慎考虑可得到的数据之后进行选择。最终选取货邮吞吐量、旅客吞吐量、起降架次、日均起降架次和日高峰架次以及平均出港滑行时间、机场跑道数量和 GDP 作为变量。

在建立体系上，我们拟选取层次分析法和熵权法来计算各机场对我国航班延误状况的综合权重，并根据权重来进行延误情况排名。

层次分析法是将决策问题按总目标、各层子目标、评价准则直至具体的备投方案的顺序分解为不同的层次结构，然后得用求解判断矩阵特征向量的办法，求得每一层次对各元素对上一层次某元素的优先权重，最后再加权和方法递阶归并各备择方案对总目标的最终权重，此最终权重最大者即为最优方案。这里所谓“优先权重”是一种相对的量度，它表明各备择方案在某一特点的评价准则或子目标，标下优越程度的相对量度，以及各子目标对上一层目标而言重要程度的相对量度。层次分析法比较适合于具有分层交错评价指标的目标系统，而且目标值又难于定量描述的决策问题。

层次分析法考虑了专家的知识经验和决策者的意向和偏好，虽然指标权重的排序往往具有较高的合理性，但仍然无法克服主观随意性较大的缺陷；熵权法充分挖掘了原始数据本身蕴涵的信息，结果比较客观，但却不能反映专家的知识经验和决策者的意向和偏好，因此，我们在模型的构建过程中，综合地运用了两种方法，降低了使用层次分析法所带来的主观随意性的问题。

2.2 对问题二

问题二主要研究的是我国航班延误的主要原因。经查阅,我们发现,在实证问题研究中,为了全面、系统地分析问题,我们必须考虑众多影响因素,如航班排队延误、航空公司延误和旅客延误等。其中又可按照可控因素和不可控因素划分得到二十多小类,分别为流量控制、军事活动和天气等不可控因素,安检、旅客、机械故障、机场原因和公共安全等可控因素。由于根据民航总局 2005—2013 年航班延误统计数据,造成中国航班延误的关键因素包括流量控制、航空公司、天气、军事活动、机场因素、机械故障以及旅客等因素,故为了数据收集的有效,将变量设置减少为流量控制、航空公司、天气、军事活动、机场因素、机械故障以及旅客。同时,经数据实证验证,各原因中并非独立,存在较为不确定的相关性和因果性,所以考虑到变量的特征,尝试运用贝叶斯网络识别航班延误的各因素间的影响程度,用于寻找航班延误的主要原因。但是,通过对各机场 8 月数据的实时监测,官方难以给我较为清晰的延误原因,且存在数据缺失和数据抽样时间段较为集中的问题。经过对数据多次的处理与分析,本文选取了较完善的数据,将变量更改为更易量化的基础变量,例如货邮吞吐量、旅客吞吐量以及起降架次和平均出港滑行时间、日均起降架次、日高峰架次以及放行正常率和 GDP。

应问题对所选变量的要求,首先对数据考虑自变量和因变量的相关性分析,但自变量间相关性的影响,容易增加误差。因此,我们选择变量建立模型,采用 R 语言编程做最优子集回归。但由于收集数据变量的误差,此方法得到的影响因素过于具体化,会造成分析结果的偏差。尽管如此,我们仍可得出收集的变量对因变量具有一定相关性的结论。考虑到每个变量都在不同程度上反映了所研究问题的某些信息,并且指标之间彼此有一定的相关性,所得的统计数据反映的信息在一定程度上有所重叠。而变量间较强的相关性会对相关性分析的研究结果和变量的选择产生较大的影响,所以本文将运用 SPSS,通过主成分分析法进行多个机场间的横向对比,设法用较少的变量来解释大部分变量,将许多相关性很高的变量转换成彼此相互独立或不相关的变量。即通过数学变换的方法,把一组相关变量通过线性变换转换成另一组不相关的变量,将这些新的变量按照方差依次递减的顺序排列,并保持变量的总方差不变。这样就能够找到影响变量方差最大的指标,即航班延误的主要原因。

由于主成分分析在主成分个数和名称确定时存在一定的主观因素影响,所以将同时运用年份间的纵向对比和官方统计数据所得影响原因验证分析结果的合理性和可靠性。最终确定该问题中的航班延误的主要原因。

### 2.3 对问题三

飞机的整个行程包括:准备起飞、飞离机场、飞离终端区、加入航线、到达终端区以及进近管制和到达机场。为提高准确率,我们考虑从优化终端区到场离场航班流的排序着手,在不违反航班安全间隔的情况下,建立模型及算法来合理安排飞机的起飞、着陆次序,以求保证交通流量的快速、有序的流动以及减少延误时间和经济损失。

目前,中国在进行航班调度上基本采用先到先服务的算法(FCFS),虽然对于管制员比较容易实施和操作,但是调度的效率低。而当航班为  $n$  架次时,其下降顺序全排列的情况有  $n!$  种,因此,当航班架次较多时,计算量非常大,且由于管制设备与水平等原因,在实际中对航班的位置进行大规模的调整亦不现实。因此,全排列中的很多排列情况并不需要考虑。

在进一步研究中发现由于机场的规模、跑道等固定不变,因此,各个航班的出发与起飞相互影响,不同的调配方式会产生不同的结果,形成“群体智能”。所谓群体智能是指“无智能的主体通过合作表现出智能行为的特征”,其中的群体是指“一组互相之间可以进行直接通信或者间接通信(通过改变局部环境),并且能够合作进行分布问题求解的主体”。

基于以上,我们拟采用提出时间较早,已得到充分发展的蚁群算法来研究问题。蚁群算

法是一种仿生算法，具有分布式、正反馈机制和贪婪式搜索的优点，具有很强的搜索能力，不易陷入局部最优解，并能很快得到全局最优解。

1) 蚁群算法是一种本质上并行的算法。每只蚂蚁搜索的过程彼此独立，仅通过信息激素进行通信。所以蚁群算法则可以看作是一个分布式的多 agent 系统，它在问题空间的多点同时开始进行独立的解搜索，不仅增加了算法的可靠性，也使得算法具有较强的全局搜索能力。

2) 蚁群算法是一种正反馈的算法。从真实蚂蚁的觅食过程中我们不难看出，蚂蚁能够最终找到最短路径，直接依赖于最短路径上信息激素的堆积，而信息激素的堆积却是一个正反馈的过程。

3) 蚁群算法具有较强的鲁棒性。

### 三、模型假设

1. 假设本文中所应用的样本 2013、2014 年主要机场的航班延误有关数据可以代表我国航班延误水平。

2. 假设官方数据及国内使用广泛的商业网站“飞常准”数据来源和分析结果可靠。

3. 假设天气原因可基本概括气象问题所引起的航班延误原因，且作为我国航班延误的主要原因。

4. 假设对航班研究原因和改进规划过程中，不会被不可控事件影响，且过程中不考虑季节因素对各变量的影响。

### 四、模型的建立与求解

4.1 对问题一：

4.1.1 根据问题分析，我们的思路如下：



图 4.1 问题一思路

#### 4.1.2 猜想一

##### 4.1.2.1 假设 2013-2015 年统计方式未改变

我们选取 2015 年 5 月 27 日深圳机场出发的航班, 分别就国内应用非常广泛的商业网站“飞常准”“flightstats.com”所给计划出发时间、实际出发时间进行对比:

以两个航班为例:

CZ6758:



图 4.2 “飞常准”和 flightstats.com 所给 CZ6758 航班相关时间

HU7014:



图 4.3 “飞常准”和 flightstats.com 所给 HU7014 航班相关时间

Flightstats.com 关于计算延误说明:

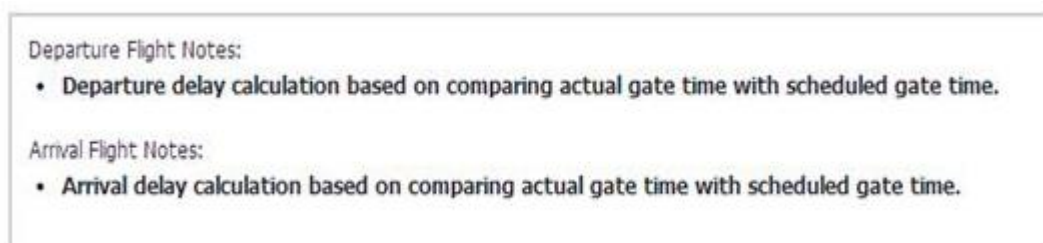


图 4.4 flightstats.com 延误说明

注: 起飞航班说明: 起飞延误统计基于比较实际关舱时间和计划关舱时间

到达航班说明: 到达延误统计基于比较实际开舱时间和计划开舱时间

经查验多个航班知, 网站所计实际出发时间均早于“飞常准”所计实际出发时间, 且相差至少 10 分钟, 因此我们可以断定 flightstats.com 的统计报告至少在 2015 年并未

采取双重标准，即未采用时间点 C-A 来计算。因此，在假设 2013-2015 年未改变统计方式的情况下，我们可以断定网站没有双重标准。

4.1.2.2 假设 2013-2015 年改变了统计方式

符号说明：

$x_i(i=1,2,\cdots 12)$ ：2013 年 1-12 月 flightstats.com 统计的 0-15min 准点率

$y_i(i=1,2,\cdots 12)$ ：2013 年 1-12 月 flightstats.com 统计的 15-30min 略晚点率

$z_i(i=1,2,\cdots 12)$ ：假设 30 分钟以内起飞为准点的“准点率”

$\eta$ ：2013 年机场平均“准点率”

$a_j(j=1,2,\cdots 12)$ ：2013 年 1-12 月 flightstats.com 统计的机场航班样本量

(1) 对 2013 年数据处理

$$z_i = x_i + y_i \tag{1}$$

处理结果如表：

表 4.1 flightstats.com 亚洲主要机场放行“准点率”

机场	1月准点率	2月准点率	3月准点率	4月准点率	5月准点率	6月准点率	7月准点率	8月准点率	9月准点率	10月准点率	11月准点率	12月准点率
Beijing (PEK)	61.56%	53.77%	60.26%	57.23%	54.84%	41.22%	41.90%	60.20%	70.06%	72.85%	78.40%	83.52%
Changsha (CSX)	63.90%	56.70%	64.63%	64.22%	65.97%	64.74%	54.85%	61.40%	72.72%	77.67%	72.20%	74.33%
Chengdu (CTU)	59.35%	59.55%	57.79%	62.32%	59.52%	54.39%	48.76%	44.85%	54.62%	71.65%	73.89%	76.61%
Chongqing (CKG)	71.48%	61.67%	72.84%	66.62%	64.94%	66.47%	54.63%	59.23%	71.31%	72.29%	71.18%	72.48%
Dalian (DLC)	79.83%	71.22%	77.77%	72.99%	72.01%	65.46%	44.00%	56.11%	75.57%	78.45%	78.65%	83.13%
Guangzhou (CAN)	65.41%	50.50%	51.23%	48.70%	55.58%	54.35%	42.48%	45.85%	56.52%	69.63%	70.26%	79.10%
Haikou (HAK)	63.50%	46.82%	63.65%	62.84%	66.88%	66.54%	57.92%	59.80%	70.77%	77.99%	77.63%	76.81%
Hangzhou (HGH)	61.03%	55.44%	59.98%	44.93%	51.34%	42.28%	28.06%	41.48%	55.80%	57.57%	64.09%	73.10%
Hong Kong (HKG)	88.38%	86.67%	82.27%	82.21%	81.18%	79.41%	74.94%	71.40%	82.98%	86.66%	82.21%	84.14%
Kunming (KMG)	63.37%	70.49%	75.86%	68.53%	71.99%	69.87%	61.06%	66.23%	78.71%	80.03%	77.31%	69.94%
Nanking (NKG)	60.19%	57.71%	61.55%	57.64%	53.37%	48.30%	34.53%	39.09%	52.86%	59.45%	53.98%	59.08%
Qingdao (TAO)	70.57%	68.83%	69.54%	96.15%	96.37%	95.39%	92.95%	94.20%	95.20%	95.84%	95.82%	94.55%
Sanya (SYX)	27.49%	19.01%	36.72%	68.15%	73.47%	66.82%	51.23%	61.35%	70.99%	76.54%	77.26%	82.95%
Shanghai (PVG)	67.82%	53.88%	62.22%	91.13%	91.38%	89.71%	81.08%	82.70%	88.66%	90.45%	90.12%	86.28%
Shanghai (SHA)	63.07%	58.84%	57.43%	59.01%	63.33%	52.27%	42.44%	54.38%	66.89%	70.55%	72.36%	76.33%
Shenyang (SHE)	65.92%	63.17%	64.39%	55.31%	57.09%	46.42%	35.08%	49.28%	70.08%	61.74%	73.90%	76.04%
Shenzhen (SZX)	69.17%	60.50%	53.75%	44.45%	48.25%	48.57%	39.43%	46.40%	58.26%	68.92%	67.34%	75.41%
Urumqi (URC)	55.90%	68.00%	78.81%	76.65%	76.88%	66.34%	58.17%	61.51%	60.89%	75.67%	67.21%	68.91%
Wuhan (WUH)	72.62%	59.46%	72.32%	70.78%	71.09%	69.42%	54.80%	65.29%	72.41%	78.72%	78.92%	81.19%
Xiamen (XMN)	69.10%	60.74%	66.27%	60.17%	53.31%	48.69%	38.45%	48.16%	54.92%	60.72%	59.75%	73.82%
Xian (XIY)	77.51%	65.33%	71.79%	69.19%	70.56%	66.23%	51.68%	65.65%	79.49%	83.48%	82.91%	85.40%
Zhengzhou (CGO)	56.37%	51.52%	63.33%	60.99%	59.38%	57.00%	39.67%	49.07%	60.25%	64.51%	56.71%	60.15%

根据其样本量加权平均得到各机场年“准点率”：

$$\eta = \frac{\sum_{i=1}^{12} z_i a_i}{\sum_{i=1}^{12} a_i} \tag{2}$$

得到结果如下表：



表 4.2 中国机场 2013 年年平均“准点率”

机场		年准点率	机场		年准点率
Beijing	(PEK)	61.43%	Nanking	(NKG)	52.41%
Changsha	(CSX)	66.41%	Qingdao	(TAO)	89.98%
Chengdu	(CTU)	60.35%	Sanya	(SYX)	61.23%
Chongqing	(CKG)	66.93%	Shanghai	(SHA)	80.67%
Dalian	(DLC)	70.47%	Shanghai	(SHA)	60.89%
Guangzhou	(CAN)	57.71%	Shengyang	(SHE)	60.01%
Haikou	(HAK)	65.76%	Shenzhen	(SZX)	56.88%
Hongkong	(HKG)	52.35%	Urumqi	(URC)	67.50%
Hangzhou	(HGH)	81.80%	Wuhan	(WUH)	70.62%
Kunming	(KMG)	71.24%	Xiamen	(XMN)	57.35%

(2) 根据国内民航局发布的《2013 年全国民航航班运行报告》能够得到几个机场的年平均准点率数据，进行数据对比得到：

表 4.3 中国机场国内外年平均准点率

机场	外国统计	中国统计
Beijing (PEK)	61.43%	66.33%
Chengdu (CTU)	60.35%	87.39%
Guangzhou (CAN)	57.71%	73.88%
Haikou (HAK)	65.76%	67.14%
Nanking (NKG)	52.41%	78.85%
Qingdao (TAO)	89.98%	83.20%
Sanya (SYX)	61.23%	42.34%
Shanghai (PVG)	80.67%	82.40%
Shanghai (SHA)	60.89%	78.86%
Shenzhen (SZX)	56.88%	68.74%
Urumqi (URC)	67.50%	85.58%

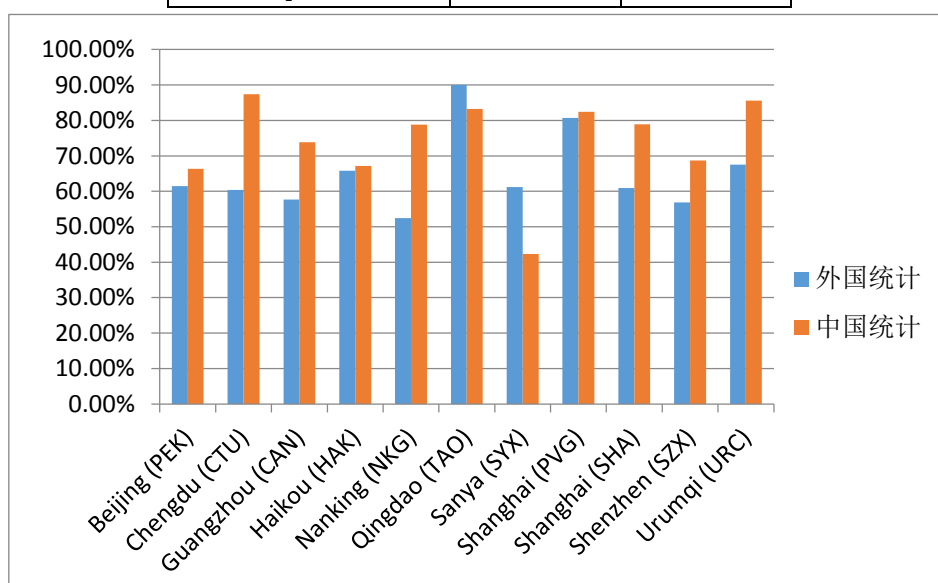


图 4.5 中国机场国内外年平均准点率



- 1) 6 个国外“准点率”明显低于国内准点率;
- 2) 2 个国外“准点率”高于国内准点率;
- 3) 3 个国外“准点率”与国内准点率接近。

### (3) 结论:

由之前的问题分析, 若未采用双重标准, 则国外“准点率”应高于国内准点率。若采用了双重标准 (即对国内的航班以 C-A 计算), 那么国外和国内的统计方式一致, 两者准点率应接近。

因此, 我们并不能说明国外采取了双重标准, 却也无法证明国外未采用双重标准。在此背景下, 我们怀疑, 出现这种情况的原因并非是双重标准的问题, 而是因为数据的处理存在着问题。

#### 4.1.3 猜想二: 数据处理出现问题

Flightstats 是以 B-A 计算的延迟时间。但是由于航空公司等原因, 并未拿到准确的实际关舱时间数据, 而是根据实际起飞时间来推测的实际关舱时间, 那么, flightstats 估计的滑行时间越短 (短于实际滑行时间), 则越可能算为延误统计量:

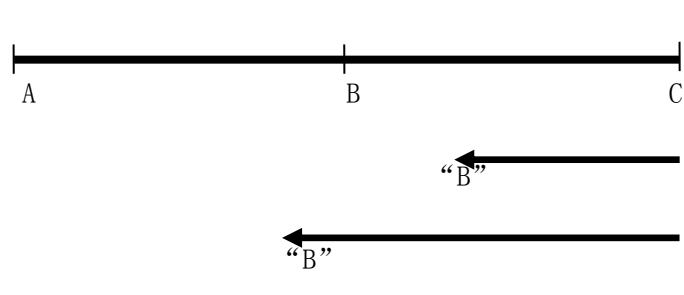


图 4.6 估算滑行时间示意

即如上图, 假设, A 到 B 用时 14 分钟, B 到 C 用时 30 分钟, 即 A 到 C 实际用时 44 分钟, 如果估计的“B”到 C 用时 20 分钟, 则 A 到“B”用时 24 分钟, 那么该航班按照 15min 以内为正点的话, 实际上是正点, 而估计则属于延误。同理, 按照 30min 来划分时段的话, 也会出现延误统计量高于实际延误统计量的情况。类似的, 也会出现延误统计量低于实际统计量的情况。因此会出现“准点率” > 准点率, “准点率” ≈ 准点率, “准点率” < 准点率, 这三种情况。进一步分析如下:

若“B” C-BC < 0, 则“准点率” < 准点率

若“B” C-BC ≥ 0, 则“准点率” > 准点率

且“B” C 与 BC 差值越大, “准点率”与准点率差值越大。

进一步以数据进行验证, 我们于《2013 年全国民航航班运行报告》查找到, 根据航空器 ACARS 计算出北京首都机场从关舱门 A 到起飞 C 平均用时 33.14min, 上海浦东机场从 A 到 C 平均用时 32.73min。而根据 EUROCONTROL 计算北京首都机场从 A 到 C 平均用时 28min, 上海浦东机场平均用时 24min。即:

对于北京首都机场: “B” C-BC = -5.14min, “准点率” - 准点率 = -4.9%

对于上海浦东机场: “B” C-BC = -8.73min, “准点率” - 准点率 = -17.97%

满足以上猜想。由于未找到 13 年其它航班的国内外航班滑行时间对比数据, 因此我们只能说该猜测具有很大可能性。

4.1.4 结论

由上述分析得出 flightstats.com 的排名并不正确。有很大可能性是在数据处理时出现了问题，导致延误率很高。但由于有“航空物语”的那一条例子，也不排除有双重标准的可能，只是经过验证，有双重标准的可能性很小。因此，虽然中国机场的延误率并不低，但是也没有 flightstats.com 统计的那么低。且 flightstats.com 公布的数据仅是 13 年 6 月份的，由于气候等问题，并不能代表中国机场的情况。而据我们查找资料的知，在 13 年 7 月 18 日中国“八大机场”不限飞后，“八大机场”的正常率有明显提高。

4.1.5 基于熵权法+AHP 构建新的评价体系

由前面分析可知，由于国内外对延误的定义不同，统计方法也不同，因此，直观的以延误率为判定进行的排名并不能另国内外均信服。因此，为了避开对延误的判定以及统计方法不同这一问题，我们不以延误率而以影响航班延误的因素作为指标构建新的评价体系。

4.1.5.1 层次分析法的实现

(1) 指标选取和层次构建

根据上文提到的构建新评价体系的原因，仔细考虑了数据的可得性以及根据问题二主成分分析和中国民航局公布的年报中的影响航班延误的因素，我们选取流量因素、机场因素和其它因素作为指标，选取货邮吞吐量、旅客吞吐量、起降架次、日均起降架次、日高峰起降架次和平均出港滑行时间、机场跑道数量、GDP 以及天气作为变量。对机场的延误状况作排名，即将机场作为备选方案，计算出机场对总目标（我国航班延误状况）的权重，根据权重进行排名即可。因此构建层次如下：

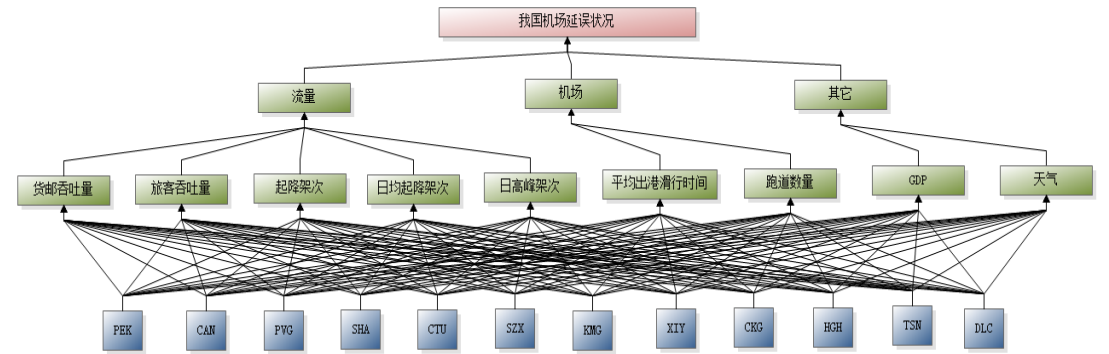


图 4.7 层次结构

(2) 判断矩阵

我们用 yaahp 软件输入数据构建判断矩阵并进行一致性检验。  
在采用 1-9 标度法时，根据中国民航局发布的年报中各因素影响延误率所占比重以及问题二得到的相对重要度进行谨慎判断且经过计算均通过了一致性检验，由于篇幅有限，我们仅在文中列举一部分，完整部分附于附录中。

表 4.4 我国机场延误状况 一致性比例：0.0836；对“我国国际产延误状况的权重：1.0000； $\lambda_{max}$ :3.0869

我国机场延误状况	流量	机场	其他	Wi
流量	1	3	0.25	0.2311
机场	0.3333	1	0.2	0.1038
其他	4	5	1	0.6651

表 4.5 流量 一致性比例: 0.0286; 对“我国机场延误状况”的权重: 0.2311;  $\lambda_{\max}$ : 5.1283

流量	货邮吞吐量	旅客吞吐量	起降架次	日均起降架次	日高峰架次	Wi
货邮吞吐量	1	0.3333	0.2	0.2	0.1429	0.0436
旅客吞吐量	3	1	0.3333	0.3333	0.2	0.0886
起降架次	5	3	1	1	0.3333	0.2017
日均起降架次	5	3	1	1	0.3333	0.2017
日高峰架次	7	5	3	3	1	0.4641

表 4.6 机场 一致性比例: 0.0000; 对“我国机场延误状况”的权重: 0.1038;  $\lambda_{\max}$ : 2.0000

机场	平均出港滑行时间	跑到数量	Wi
平均出港滑行时间	1	0.3333	0.25
跑到数量	3	1	0.75

表 4.7 其他 一致性比例: 0.0000; 对“我国机场延误状况”的权重: 0.6651;  $\lambda_{\max}$ : 2.0000

机场	平均出港滑行时间	跑到数量	Wi
GDP	1	0.1111	0.1
天气	9	1	0.9

(3) 层次分析法权重的得出  
我们利用 yaahp 软件计算权重结果如下:

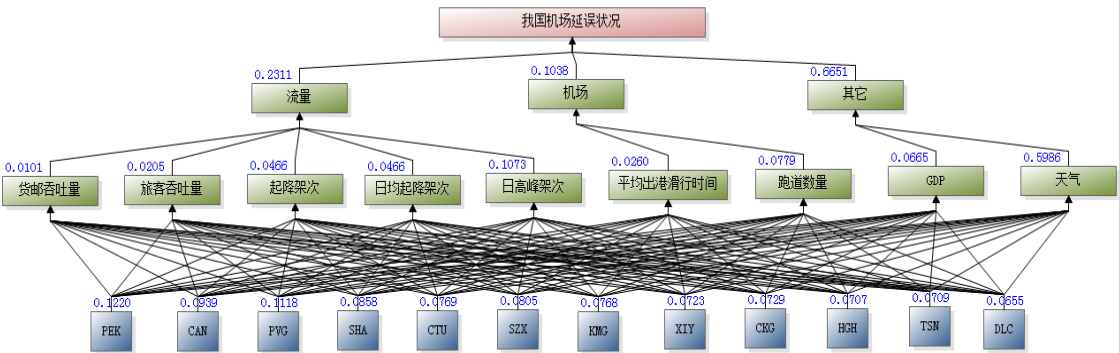


图 4.8 AHP 权重计算结果

4.1.5.2 熵权法的实现

一般来说, 若某个指标的信息熵越小,  $E_j$  表明指标值得变异程度越大, 提供的信息量越多, 在综合评价中所能起到的作用也越大, 其权重也就越大。相反, 某个指标的信息熵  $E_j$

越大，表明指标值得变异程度越小，提供的信息量也越少，在综合评价中所起到的作用也越小，其权重也就越小。

首先，将各个指标的数据进行标准化处理。假设给定了  $k$  个指标  $X_1, X_2, \dots, X_k$ ，其中  $X_i = \{x_1, x_2, \dots, x_k\}$ ，假设对各个指标进行标准化后的值为  $Y_1, Y_2, \dots, Y_k$ ，那么，

$$Y_{ij} = \frac{x_{ij} - \min(x_i)}{\max(x_i) - \min(x_i)} \tag{3}$$

接下来，根据信息论中信息熵的定义，一组数据的信息熵

$$E_j = -\ln(n)^{-1} \sum_{i=1}^n p_{ij} \ln p_{ij} \tag{4}$$

其中  $p_{ij} = Y_{ij} / \sum_{i=1}^n Y_{ij}$ ，如果  $p_{ij} = 0$ ，则定义

$$\lim_{p_{ij} \rightarrow 0} p_{ij} \ln p_{ij} = 0 \tag{5}$$

根据信息熵的计算公式，计算出各个指标的信息熵： $E_1, E_2, \dots, E_k$ 。通过信息熵计算出各个指标的权重

$$W_i = \frac{1 - E_j}{k - \sum E_j} (I = 1, 2, \dots, k) \tag{6}$$

计算结果如下表：

表 4.8 熵权法权重计算结果

	PEK	CAN	PVG	SHA	CTU	SZX	KMG	XIY	CKG	HGH	TSN	DLC
W'	0.083	0.082	0.078	0.085	0.084	0.082	0.085	0.086	0.085	0.084	0.083	0.085

4. 1. 5. 3 综合权重的得出  
 本文把 AHP 法和熵权法的结果相结合, 得到综合考虑主客观因素的指标权重向量  $\omega = (\omega_j)_{1 \times n}$ ，即

$$\omega_j = \alpha \omega_j' + (1 - \alpha) \omega_j'' \quad 0 \leq \alpha \leq 1 \tag{7}$$

综合权重随  $\alpha$  的改变而改变, 当  $\alpha = 1$  和  $\alpha = 0$  时, 分别对应于 AHP 法和熵权法  $\alpha$  如何合理地取值, 有很多讨论，我们根据各指标的重要程度等级排序、AHP 所得指标权重排序和熵权法所得指标权重排序, 根据它们的一致程度最终选取  $\alpha = 0.5$ 。

综合权重计算结果如下：

表 4.9 综合权重

	PEK	CAN	PVG	SHA	CTU	SZX	KMG	XIY	CKG	HGH	TSN	DLC
w	0.103	0.088	0.095	0.085	0.08	0.081	0.081	0.079	0.079	0.077	0.077	0.075

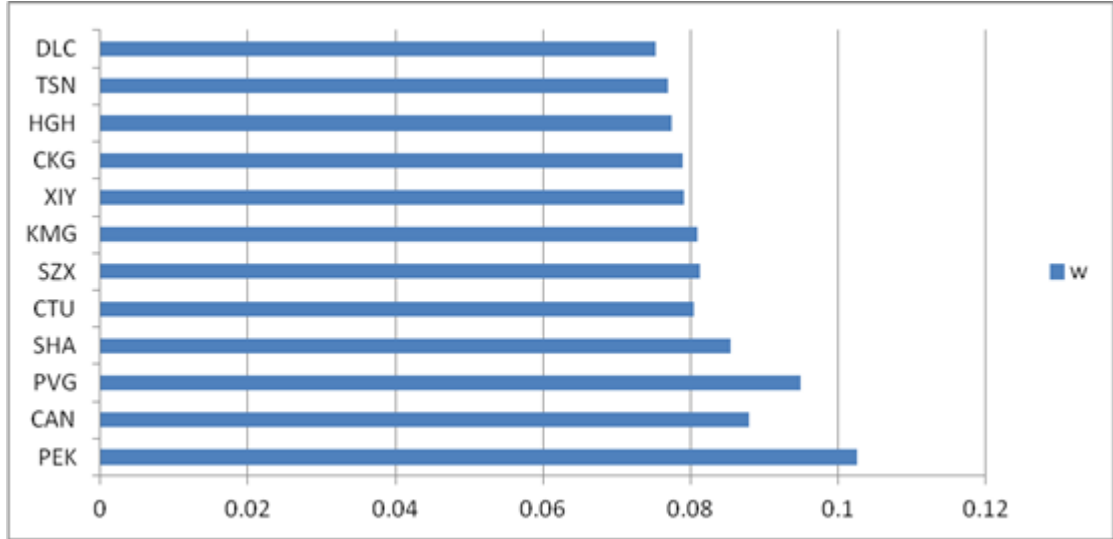


图 4.9 根据综合权重排名

无论是国内还是国外网站统计的延误率，北京首都机场和上海浦东机场的延误情况都很严重，在新建立的评价体系中我们亦可看出，北京首都机场和上海浦东机场的延误情况最严重，说明，该评价体系具有合理性。

4.2 对问题二

4.2.1 解题思路：

(1) 基于民航总局 2005—2013 年航班延误统计数据，将变量初步确定为流量控制、航空公司、天气、军事活动、机场因素、机械故障以及旅客因素，并对变量间进行相关性检验，通过检验结果讨论建模方法。

(2) 检验结果显示各变量存在相关性，为非独立的变量，故针对变量间的关系不确定性，本文将运用贝叶斯网络通过数据对各变量的影响程度进行训练和检验，最终得到航班延误的主要原因。

(3) 通过对各大机场 8 月航班数据的实时监测和收集，发现官方未能及时提供航班延误的主要原因，存在严重的数据缺失和数据抽样时间段较集中的问题。故本文经过对数据多次的处理、对比与分析，决定将原有变量替换为基础的量化数据，例如货邮吞吐量、旅客吞吐量以及起降架次和平均出港滑行时间、日均起降架次、日高峰架次以及放行正常率和 GDP。最终对这些数据进行通过最优子集回归验证变量提取的可行性并对变量进行相关性检验，由于变量间仍然存在相关性，故应用主成分建模分析。

(4) 在对原变量的替换过程中，通过对天气因素和天气间接影响因素的研究发现，天气因素的多数基础指标较为抽象，难以量化研究或是量化影响因素过于繁杂，使得误差过大。但是，若出现雷暴等特殊天气，将直接影响航班延迟或取消，间接影响航线占用、航道规划等变量，因此难以忽略天气原因的重要性，故直接将天气因素归因于出现航班延误的现象的主要原因。

(5) 由于主成分结论的强主观性，则需要多角度的验证，用于提高结论的现实意义。所以，本文将分别从纵向和官方数据比较等方面进行模型的验证。

a、纵向来看，从不同年份的变量变化，也可看出，与横向分析结果基本无差别，以此

证明其合理性。

b、官方数据对比，根据分析结果可验证横向对比的可靠性。

4.2.2 变量选择

4.2.2.1 初步选择

根据民航总局 2005—2013 年航班延误统计数据，造成中国航班延误的关键因素包括流量控制、航空公司、天气、军事活动、机场因素、机械故障以及旅客等因素，综合主要航空公司和主要机场的相关数据，得到各因素引发的延误比例结构。

表 4.10 各因素引发的延误比例

年份	流量控制	航空公司	天气	军事活动	机场	旅客
2005	0.16	0.46	0.18	0.03	0.04	0.03
2006	0.22	0.48	0.23	0.02	0.02	0.02
2007	0.28	0.47	0.15	0.07	0.02	0.01
2008	0.19	0.43	0.27	0.06	0.04	0.01
2009	0.23	0.39	0.19	0.11	0.05	0.03
2010	0.24	0.41	0.23	0.09	0.02	0.01
2011	0.28	0.37	0.20	0.12	0.01	0.02
2012	0.22	0.36	0.21	0.17	0.02	0.02

这些统计数据仅反映了各关键影响因素引起航班延误发生频率的比重，而没有体现对航班延误。但是，由上表可将影响因素减少至流量控制、航空公司、天气、军事活动、机场因素、机械故障以及旅客因素。由于选取变量间存在相关性和因果性，故将选用贝叶斯网络通过数据对各变量的影响程度进行训练和检验，最终得到航班延误的主要原因。

然而，通过对各大机场 8 月航班数据的实时监测和收集，发现官方未能及时提供航班延误的主要原因，存在严重的数据缺失和数据抽样时间段较集中的问题。导致贝叶斯网络中的结点过少，没有实际意义。所以，本文通过对比分析决定进行变量替换。

4.2.2.2 变量替换

机场（2014 年）：由于我国机场数量较多，非主要繁忙机场容易出现统计数据缺失量较大的情况，对统计分析结果影响较大。因此本文选用我国主要繁忙机场估计我国机场的总体情况。并且，出于对数据精准性和实效性的考虑，本文选取了《2014 年全国民航航班运行效率报告》中的数据作为模型处理数据。

主要繁忙机场放行正常率：又称为准点率，作为衡量机场准点概率的主要衡量因素。

货邮吞吐量（吨）：报告期内经由航空进、出港区范围的货邮数量。

旅客吞吐量（人）：报告期内经由航空进、出港区范围的旅客数量。

起降架次（次）：是指一年内航班起降的总次数，与飞机数量没有关系，一架飞机在一个较长的单位时间内可以累计多个起降架次，计算方式类同于参观人次或者发送车次等。

平均出港滑行时间（分钟）：是指出港航班从跑道到到停机位的平均滑行时间。

日交通量（日均起降架次）：是指一年内每日平均航班起降的次数。

日交通量（日高峰架次）：是指一年内每日航班起降的最高次数。

GDP：国内主要繁荣机场所在城市不是政治、经济中心，就是各区域的交通枢纽中心，旅游集散地。因民航运输需求与经济、人口、旅游以及所处地理位置等几个因素相关，因此 GDP 可作为民航业发展的衡量指标之一。

### 4.2.3 数据收集

以下数据均来自《2014 年全国民航航班运行效率报告》中对于主要繁忙机场的数据研究。

表 4.11 2014 年中国主要繁忙机场数据

X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	X <sub>7</sub>	X <sub>8</sub>	X <sub>9</sub>
北京首都	69.69	1848251.48	86128313	581952	21.6	1594	1816	21330.8
广州白云	68.82	1454043.78	54780346	412210	19	1130	1309	16706.87
上海浦东	56.25	3181654.06	51687894	402105	22	1102	1365	23560.94
上海虹桥	67.72	432176.398	37971135	253325	17	696	847	23560.94
成都双流	82.27	545011.232	37675232	270054	14.5	755	832	10056.6
深圳宝安	67.32	963871.182	36272701	286346	17.5	784	934	16001.98
昆明长水	81.83	316672.439	32230883	270529	16	743	837	3712.99
重庆江北	78.28	302335.812	29264363	238085	17.9	653	739	14265.4
西安咸阳	80.1	186412.614	29260755	245971	18.5	679	768	5474.77
杭州萧山	61.4	398557.606	25525862	213268	17.3	585	693	9201.16
厦门高崎	60.38	306384.994	20863786	174315	21	484	535	3273.54
长沙黄花	70.57	125037.764	18020501	152359	20	418	516	7824.81
武汉天河	75.09	143029.57	17277104	157596	17.3	432	555	10069.48
青岛流亭	68.19	204419.401	16411789	142452	15.3	390	475	8692.1
乌鲁木齐 地窝堡	81.37	162711.318	16311140	142266	19.2	389	455	2510
南京禄口	48.35	304324.819	16283816	144278	17.7	396	466	8820.75
三亚凤凰	56.41	75645.842	14942356	102074	15.4	293	402	404.38
海口美兰	58.74	121131.4	13853859	105861	16	290	482	1005.51
大连 周水子	73.73	133490.023	13551223	115284	12.7	317	431	7655.6
天津滨海	58.7	233358.644	12073041	114557	19.1	315	465	15722.5

注：X<sub>1</sub>：机场（2014 年），X<sub>2</sub>：主要繁忙机场放行正常率 %，X<sub>3</sub>：货邮吞吐量（吨），X<sub>4</sub>：旅客吞吐量（人），X<sub>5</sub>：起降架次（次），X<sub>6</sub>：平均出港滑行时间（分钟），X<sub>7</sub>：日均起降架次，X<sub>8</sub>：日高峰架次，X<sub>9</sub>：GDP（亿元）。

### 4.2.4 建立模型

#### 4.2.4.1 最优子集回归（best-subset selection）

##### （1）方法介绍

Subset Selection 是一个可能的降低 Variance 的方式，事实上由于可能样本数据中的维数比较高，学习方法会比较倾向于过拟合数据（Overfitting），Subset Selection 则倾向于选择较少的和输出相关度最高的维度，并利用这些维度进行 Regression。另外一个好处则是能更好的解释数据，通过寻找最相关的几个特征，我们能更好地直观感受输出结果受那几个输入影响最大。

Best-Subset Selection 试图寻找全部特征的一个子集，并利用这个子集进行 Least Square Regression，通过限制子集的大小，我们可以控制算法的 Bias-Variance Tradeoff。子集选择的过程需要一点技巧，因为子集的个数是指数相关于子集大小的，我们利用一点迭代技巧来处理这个问题。我们可以选择从一个空的子集中逐步加入和目标函数相关系数最大



的特征（或者说加入这个特征可以使得 Squared Loss 减少最多），同样也可以选择从满的特征集合中逐步 drop 相关性最小的特征，这里用到了统计学中的 Z-score，我们不去细表。

与子集选择非常相似的一个方法则是 Stagewise Regression，这里我们每次选择一个特征，然后把它设为最优值（即使得 Squared Loss 尽量小），每次我们选择的特征应该是和目标相关度最大的特征或者调整后能使 Squared Loss 减小最多的特征。Stagewise Regression 相比于 Least Square 等方法速度要慢很多，因为它使用了一种迭代性的求解方式而不是解析性的求解，这使得在历史中这个方法并没有收到很大重视，但是最近的研究表明在高维空间中这种“Slow fitting”可以很好地降低 Variance。

Cp 准则：1964 年马勒斯（Mallows）从预测的角度提出了一个可以用来选择自变量的统计量，这就是我们常说的 Cp 统计量。

Cp 统计量为

$$C_p = \frac{SEE_p}{\hat{\sigma}^2} - n + 2p = (n - m - 1) \frac{SSE_p}{SSE_m} - n + 2p \tag{1}$$

选择变量的 Cp 准则：选择使 Cp 最小的自变量子集，这个自变量子集对应的回归方程就是“最优”回归方程。

(2) 模型建立

1) 将数据以矩阵的形式导入 R 内，使用 regsubsets 函数进行模型建立变量选择，得到分别从一个变量至七个变量的模型建立结果。

表 4.12 模型筛选结果

变量个数	货邮吞吐 量	旅客吞 吐量	起降架 次	平均出 港滑行 时间	日均起 降架次	日高峰 架次	GDP
1 ( 1 )				☆			
2 ( 1 )	☆				☆		
3 ( 1 )	☆	☆			☆		
4 ( 1 )	☆	☆		☆	☆		
5 ( 1 )	☆	☆		☆	☆	☆	
6 ( 1 )	☆	☆	☆	☆	☆	☆	
7 ( 1 )	☆	☆	☆	☆	☆	☆	☆

注释：表中“☆”表示该变量被选入，空格表示变量不被选入

2) 从 7 个备选模型中得到按 cp 准则排列的效果图，并提取得到 cp 值最小的模型作为最优模型，同时提取此模型的变量系数。

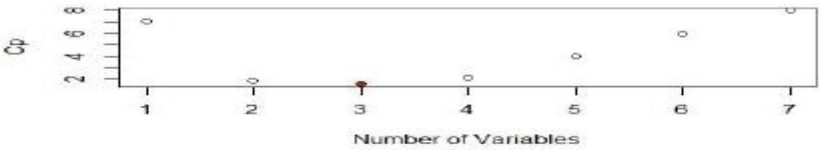


图 4.10 Cp 准则下的模型选择 (1)

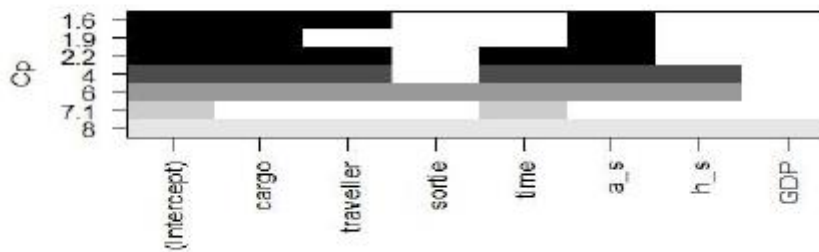


图 4.11 Cp 准则下的模型选择 (2)

表4.13 最优模型系数

截距项	货邮吞吐量	旅客吞吐量	日均起降架次
5.262084e+01	-1.302676e-05	-1.269018e-06	9.641205e-02

由此能说明选择出的变量对准点率有一定程度的线性关系,但由于由此选择出的变量所得原因过于具体,所以在证明了与准点率的相关性后,本文将运用主成分分析得到航班延误的主要原因。

#### 4.2.4.2 主成分分析法

##### (1) 方法介绍

1) 主要成分分析: 把原来多个变量划为少数几个综合指标的一种统计分析方法,是一种降维处理技术。

记原来的变量指标为, 它们的综合指标  $x_1, x_2, \dots, x_p$  ——新变量指标为

$$z_1, z_2, \dots, z_m (m \leq p), \quad (2)$$

则

$$\begin{cases} z_1 = l_{11}x_1 + l_{12}x_2 + \dots + l_{1p}x_p \\ z_2 = l_{21}x_1 + l_{22}x_2 + \dots + l_{2p}x_p \\ \vdots \\ z_m = l_{m1}x_1 + l_{m2}x_2 + \dots + l_{mp}x_p \end{cases} \quad (3)$$

$z_1, z_2, \dots, z_m$  分别称为原变量指标  $x_1, x_2, \dots, x_p$  的第一, 第二,  $\dots$ , 第  $m$  主成分, 在实际问题的分析中, 常挑选前几个最大的主成分。

1)  $z_i$  与  $z_j$  ( $i \neq j, i, j = 1, 2, \dots, m$ ) 相互无关;

2)  $z_1$  是  $x_1, x_2, \dots, x_p$  的一切线性组合中方差最大者,  $z_2$  是与  $z_1$  不相关的  $x_1, x_2, \dots, x_p$  的所有线性组合中方差最大者,  $\dots$ ,  $z_m$  是与  $z_1, z_2, \dots, z_{m-1}$  都不相关的  $x_1, x_2, \dots, x_p$  的所有线性组合中方差最大者。

找主成分就是确定原来定量  $x_j(j=1,2,\cdots,m)$  在诸主成分  $z_i(i=1,2,\cdots,m)$  上的载荷

$l_{ij}(i=1,2,\cdots,m, j=1,2,\cdots,p)$ 。它们分别是  $x_1, x_2, \cdots, x_p$  的相关矩阵的  $m$  个较大的特征值所对应的特征向量。

2) 主成分分析主要由以下几个方面的作用

a. 主成分分析能降低所研究的数据空间的维数。即用研究  $m$  维的  $Y$  空间代替  $p$  维的  $X$  空间 ( $m < p$ )，而低维的  $Y$  空间代替高维的  $x$  空间所损失的信息很少。即：使只有一个主成分  $Y_1$  (即  $m=1$ ) 时，这个  $Y_1$  仍是使用全部  $X$  变量 ( $p$  个) 得到的。例如要计算  $Y_1$  的均值也得使用全部  $x$  的均值。在所选的前  $m$  个主成分中，如果某个  $X_i$  的系数全部近似于零的话，就可以把这个  $X_i$  删除，这也是一种删除多余变量的方法。

b. 有时可通过因子负荷  $a_{ij}$  的结论，弄清  $X$  变量间的某些关系。

c. 多维数据的一种图形表示方法。我们知道当维数大于 3 时便不能画出几何图形，多元统计研究的问题大都多于 3 个变量。要把研究的问题用图形表示出来是不可能的。然而，经过主成分分析后，我们可以选取前两个主成分或其中某两个主成分，根据主成分的得分，画出  $n$  个样品在二维平面上的分布况，由图形可直观地看出各样品在主分量中的地位，进而还可以对样本进行分类处理，可以由图形发现远离大多数样本点的离群点。

d. 由主成分分析法构造回归模型。即把各主成分作为新自变量代替原来自变量  $x$  做回归分析。

e. 用主成分分析筛选回归变量。回归变量的选择有着重的实际意义，为了使模型本身易于

做结构分析、控制和预报，好从原始变量所构成的子集合中选择最佳变量，构成最佳变量集合。用主成分分析筛选变量，可以用较少的计算量来选择量，获得选择最佳变量子集合的效果。

(2) 模型建立

1) 将数据进行标准化处理后建立相关矩阵。

表 4.14 原有变量的相关系数矩阵

	货邮吞吐 量	旅客吞吐 量	起降 架次	平均出港 滑行时间	日均起 降架次	日高峰 架次	GDP
货邮吞吐量	1.000	0.773	0.798	0.572	0.797	0.835	0.704
旅客吞吐量	0.773	1.000	0.991	0.492	0.992	0.988	0.687
起降架次	0.798	0.991	1.000	0.526	0.999	0.993	0.686
平均出港滑行时间	0.572	0.492	0.526	1.000	0.522	0.529	0.419
日均起降架次	0.797	0.992	0.999	0.523	1.000	0.993	0.683
日高峰架次	0.835	0.988	0.994	0.529	0.993	1.000	0.714
GDP	0.704	0.687	0.686	0.419	0.683	0.714	1.000

可以看到，大部分相关系数都较高，各变量呈现较强的线性关系，能够从中提取主成分，适合进行主成分分析。

表 4.15 巴特利特球度检验和 KMO 检验

KMO 和 Bartlett 的检验

取样足够度的Kaiser-Meyer-Olkin度量		.791
Bartlett的球形度检验	近似卡方	347.200
	df	21
	sig.	.000

由表 4.15 可知，巴特利特球度检验统计量的观测值为 347.2，相应的概率  $p$  值接近 0。如果显著性水平为 0.05，由于概率  $p$  值小于显著性水平，应拒绝零假设，认为相关系数矩阵与单位阵有显著差异。同时，KMO 值为 0.791，根据 Kaiser 给出的 KMO 度量标准可知，原有变量适合进行主成分分析。

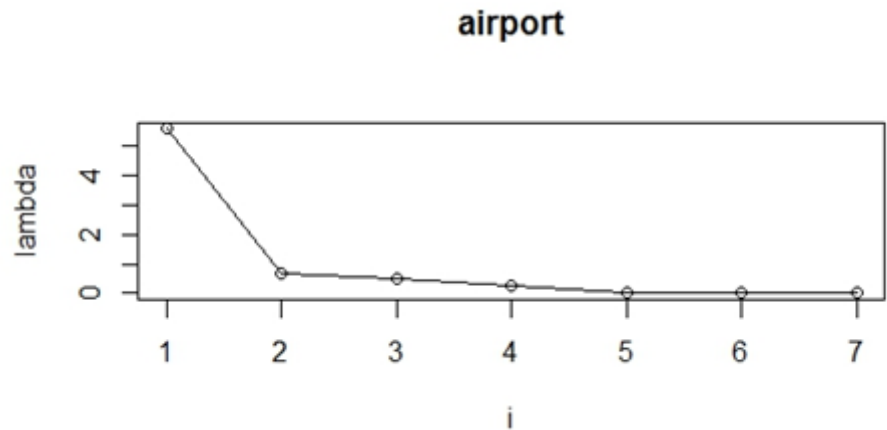


图 4.12 碎石图

由碎石图指定提取特征根的标准，指定提取两个因子。分析结果如下表所示。

表 4.16 公因子方差

	初始	提取
货邮吞吐量（吨）	1.000	.787
旅客吞吐量（人）	1.000	.965
起降架次（次）	1.000	.972
平均出港滑行时间（分钟）	1.000	.978
日交通量（日均起降架次）	1.000	.971
日交通量（日高峰架次）	1.000	.986
GDP（亿元）	1.000	.612

由表 4.16 第二列可知，此时所有变量的共同度均较高，各个变量的信息丢失都较少，因此本次提取的总体效果较理想。

表 4.17 因子解释所有变量总方差的情况

解释的总方差

成分	初始的特征根			提取平方和载入			旋转平方和载入		
	合计	方差的%	累计%	合计	方差的%	累计%	合计	方差的%	累计%
1	5.585	79.782	79.782	5.585	79.782	79.782	4.800	68.569	68.569
2	.687	9.815	89.596	.687	9.815	89.596	1.472	21.028	89.596
3	.473	6.750	96.346						
4	.241	3.444	99.790						
5	.010	.139	99.929						
6	.005	.070	99.999						
7	5.273E-5	.001	100.00						

第一组数据项描述了初始因子解的情况。在初始解中由于提取了 7 个因子，因此原有变

量的总方差均被解释掉。

第二组数据项描述了因子解的情况。可以看到，由于指定提取两个因子，两个因子共解释了原有变量总方差的 89.596%。总体上原有变量的信息丢失较少，分析效果较理想。

第三组数据项描述了最终因子解的情况。可见，因子旋转后，累计方差贡献率没有改变，也就是没有影响原有变量的共同度，但却重新分配了各个因子解释原有变量的方差，改变了各因子的方差贡献，使得因子更易于解释。

表 4.18 因子载荷矩阵

	成分	
	1	2
货邮吞吐量（吨）	.880	.114
旅客吞吐量（人）	.967	-.174
起降架次（次）	.977	-.133
平均出港滑行时间（分钟）	.622	.769
日交通量（日均起降架次）	.976	-.137
日交通量（日高峰架次）	.985	-.123
GDP（亿元）	.782	-.032

由表 4.18 可得到第一主成分为空管成分，第二主成分为机场成分。由于主成分分析中，对第一和第二主成分的成分总结中主观因素较大，因此，运用以下对问题进行不同方法和维度的解决方案验证横向对比结论的合理性和可靠性。

4.2.5 纵向对比

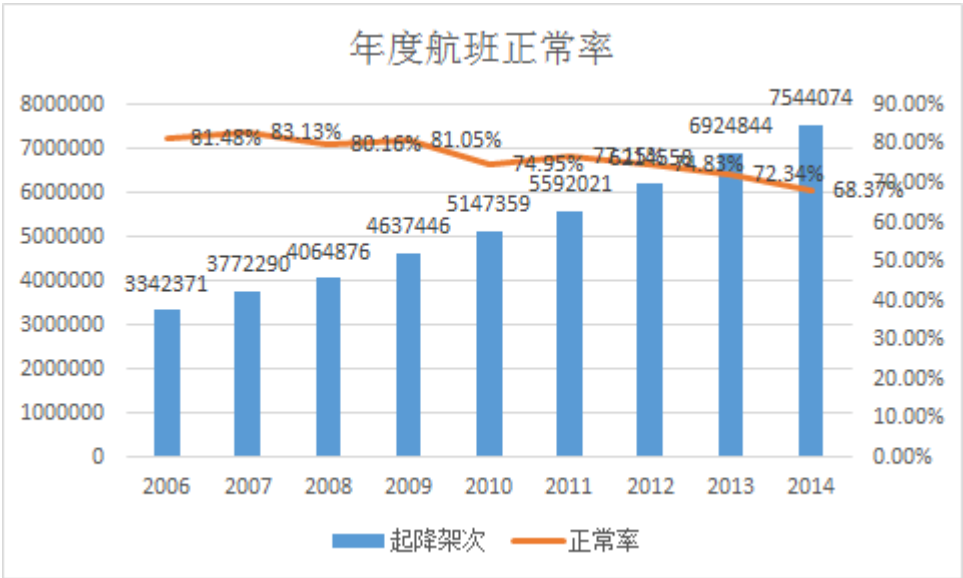


图 4.13 年度航班正常率与起飞架次对比图

由图 4.13 可见，随着起降架次的逐年提高，航班正常率显著下降。2006 年至 2014 年全国航班量持续快速增长，9 年间增长了 130%，年均增长达 10.8%，但是航班正常率也呈下滑趋势，2009 年之前正常率都在 80% 以上，2010 年开始持续下降，2014 年下降到了 68.37%，年均约下降 1.46%。同理，经由对不同变量数据的研究，可以基本证明横向对比得到的主要原因的合理性。

4.2.6 官方权威调查数据对比

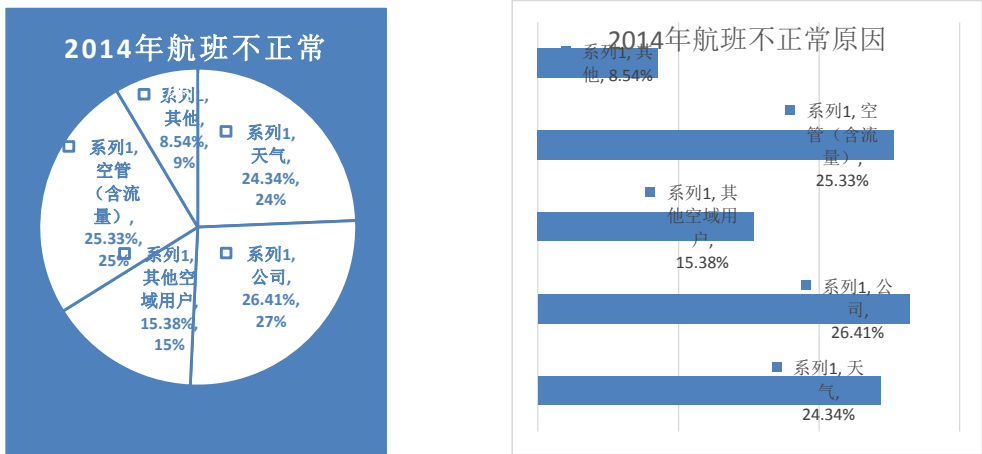


图 4.14 航班不正常原因占比

根据《2014 年全国航班运行效率报告》对年度航班不正常因素的总结中，可以得到导致航班不正常的原因主要包括天气、公司、其它空域用户、空管（含流量）、时刻安排、机场、旅客以及公共安全。如上图所示，2014 年航班延误最主要因素是公司、空管和天气原因，分别达到 26.41%、25.33%和 24.34%。此结论与横向对比基本一致，证明了横向对比结论的可靠性。

4.3 对问题三

4.3.1 符号说明：

$t$  :延误时间

$c_0$  :各机型每小时延误运营成本

$OC(t)$  :延误的运营成本

$r$  : 中国航空公司的平均净利润率

$s$  :航班客座率

$m$  :航班最大载客人数

$f$  :平均飞行时间

$p$  :各飞机每小时可为航空公司赚取的利润

$a$  : 平均票价

$UG(t)$  : 盈利损失

$TFL(t)$  :旅客经济损失

$b$  : 国内航班飞行每名旅客平均延误经济损失

4.3.2 基于蚁群算法的航班延误改进

根据民生证券 2013 年在新浪财经发布的《航班延误所致经济损失每年或达千亿元》的分析报告指出，在不考虑间接成本的情况下，我国航班延误 2013 年导致的经济损失在低延误情况下为 439 亿元，在高延误情况下高达 878 亿元，损失十分巨大。

而无论是何种原因造成的航班延误，减少航班延误损失的最直接有效方式是在尽可能短的的时间内恢复航班，使航班延误所造成的经济损失降到最少。根据这一目标，我们寻找采用蚁群算法进行优化求解。首先我们对这一目标进行构成分析。

4.3.2.1 航班延误原因

造成航班延误的因素很多，根据对问题二的分析，可知主要原因有三：

- (1) 是天气等自然因素造成的延误；
- (2) 是飞机出现机械故障、飞机调配、发行计划安排出现问题等航空公司自身的原因造成的延误；
- (3) 是空中流量饱和，空管对飞机起飞进行控制造成的延误。

种种原因造成的航班延误不仅耗费旅客的时间带来种种不便，对航空公司而言，加重了运营成本，飞机利用率和机组利用率的下降也导致了盈利损失。

4.3.2.2 航班延误损失

我们从三方面分析航班延误的经济损失，因而我们假设航班延误的经济损失构成（此处不包括名誉诚信等受损所导致的后继经济损失）如下图：

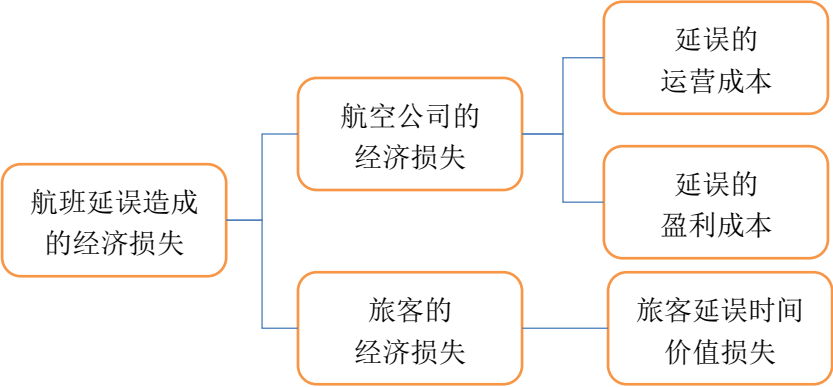


图 4.15 航班延误造成的经济损失构成

航空公司的经济损失由如下三个方面构成。

(1) 运营损失

本文参照《基于经济损失的航班延误恢复模型研究》中的假设，重型机、中型机和轻型机延误的运营成本分别为 10 万元 / 天、7 万元 / 天和 5 000 元/天，则有各机型每小时延误运营成本  $c_0$  的值如表 1 所示。

表 4.19 各类飞机每小时延误运营成本

机型	代表符号	最大起飞质量/t	运营成本 $c_0$ /元 $\cdot h^{-1}$
重型	H	>136	4167
中型	M	7-136（含）	2916
轻型	L	$\leq 7$	208

则对于不同机型其延误的运营成本  $OC(t)$ （operating costs）



$$OC(t) = c_0 t \quad (1)$$

## (2) 盈利损失

飞机是航空公司创造利润的主要工具，当航班发生延误时，会导致飞机利用率和机组利用率的下降，从而导致航空公司的盈利受到影响，造成经济损失。延误航班的盈利损失主要与航班最大载客人数、航班客座率和航空公司的净利润率有关。根据 2004 年英国《商业航空》<sup>[5]</sup>的统计可得出中国航空公司的平均净利润率  $r$  约为 2.98%。参照《基于经济损失的航班延误恢复模型研究》，根据我国国情，假设航班客座率  $s$  为 75%，平均票价  $a$  为 750 元，平均飞行时间  $f$  为 2h，航班最大载客人数由  $m$  表示，则各飞机每小时可为航空公司赚取的利润  $p$  为

$$p = \frac{m \cdot s \cdot a \cdot r}{f} \quad (2)$$

延误航班的盈利损失  $UG$  (unrealized gain) 为:

$$UG(t) = p \cdot t \quad (3)$$

## (3) 旅客的经济损失

根据《基于经济损失的航班延误恢复模型研究》中的分析，中国民航运输中对于普通国内航班飞行每名旅客平均延误经济损失  $b$  约为 50 元  $\cdot h^{-1}$ ，则航班延误造成的旅客经济损失  $TFL$  (travelers' financial loss) :

$$TFL(t) = b \cdot t \quad (4)$$

### 4.3.3 基于蚁群算法的航班延误损失优化模型的建立

由于某些原因造成了大面积的航班不能正常起飞，当延误恢复时，要重新调度这些延误了的航班。调度的任务是给出一种航班调度方案，将执行航班恢复的时间平均分成若干段，每段用  $\Delta t_j$  表示，要使得该时段内延误的总成本最小，目标函数如下:

$$\min FL = \min \sum_{j=1}^n [OC(t_j) + UG(t_j) + TFL(t_j)] \quad (5)$$

s.t.

$$\begin{cases} \sum_{i=1}^n l_{ij} \leq m \Delta t_j, \\ t_j \in [0, e \Delta t_j + d_i]. \end{cases} \quad (6)$$

其中， $e \Delta t_j$  表示航班  $f_j$  的预计降落时间所在的时间段值； $d_j$  表示航班  $f_j$  降落后允许的最大延误时间； $l_{ij} = 0$  表示航班  $f_j$  未在  $\Delta t_j$  时间段降落，等于 1 则表示航班  $f_j$  在  $\Delta t_j$  时间段

降落； $m\Delta t_j$  表示在时间段  $\Delta t_j$  的机场的容量值。

#### 4.3.3.1 蚁群算法的简介

蚁群算法 (ant colony algorithm, ACA) 是由意大利学者 M.Dorigo 等人于 20 世纪 90 年代初提出的一种新的模拟进化算法，其真实地模拟了自然界蚂蚁群体的觅食行为。

自然界中的蚂蚁觅食是一种群体性行为，蚂蚁在寻找食物源时，会在其经过的路径上释放一种信息素，并能够感知其他蚂蚁释放的信息素。信息素浓度的大小表征路径的远近，信息素浓度越高，表示对应的路径距离越短。通常，蚂蚁会以比较大的概率优先选择信息素浓度较高的路径，并释放一定量的信息素，以增强该条路径上的信息素浓度，这样会形成一个正反馈。最终，蚂蚁能够找到一条从巢穴到食物源的最佳路径，即最短距离。路径上的信息素浓度会随着时间的推进而逐渐衰减。

#### 4.3.3.2 模型建立

用传统的蚁群算法解决航班排序问题，给出一个操作时段内所有飞机预计到达时间，把每架飞机抽象为一个顶点，为任意两个顶点建立一条有向边，其中边的长度定义为两架飞机之间的降落时间间隔，则使航班排序问题的优化就等价于求解一条周游所有顶点的最短路径，从而建立起了一个 TSP 模型，再用蚁群算法对此问题优化求解。

本文中，模型建立的思想类似，不同的是，将目标函数的最短延误时间，改为最小延误经济损失，这里对延误经济损失的计算其实是一个关于延误时间的线性函数。

将蚁群算法用以解决此问题有以下步骤：

##### (1) 初始化参数

设定蚂蚁规模  $m=50$ ，航班数量为  $n$ ，航班  $i$  与航班  $j$  之间的距离为  $d_{ij}$  ( $i, j=1, 2, \dots, n$ )， $\alpha$  为信息素重要程度因子，其值越大，表示信息素的浓度在转移中的作用越大，设定  $\alpha=1$ ， $\beta$  为启发函数重要程度因子，其值越大，表示启发函数在转移中的作用越大，即蚂蚁会以较大的概率转移到尾流间隔短的航班，设定  $\beta=2$ ，信息素挥发因子  $\rho$  表示信息的挥发程度  $\rho=0.01$ ，信息素释放总量  $Q=1000$ ，最大迭代次数  $NC_{\max}=200$ ，迭代次数初值  $NC=1$ 。

##### (2) 构建解空间

将  $m$  只蚂蚁放到  $n$  个航班上，对每个蚂蚁  $k$  ( $k=1, 2, \dots, m$ )，按照式计算其下一个待访问的航班，直到所有蚂蚁访问完所有航班。

设  $P_{ij}^k(t)$  表示  $t$  时刻蚂蚁  $k$  从航班  $i$  转移到航班  $j$  的概率，其计算公式为：

$$P_{ij}^k = \begin{cases} 0, & s \notin to\_visit_k \\ \frac{[\tau_{ij}(t)]^\alpha \cdot [\eta_{ij}(t)]^\beta}{\sum_{s \in to\_visit_k} [\tau_{is}(t)]^\alpha \cdot [\eta_{is}(t)]^\beta}, & s \in to\_visit_k \end{cases} \quad (7)$$

其中

1)  $\tau_{ij}(t)$  表示  $t$  时刻航班  $i$  与航班  $j$  之间的尾流间隔的信息素浓度，初始时，各航班间的信息素浓度相同；

2)  $\eta_{ij}(t)$  为启发函数， $\eta_{ij}(t) = 1/d_{ij}$ ，表示蚂蚁从航班  $i$  转移到航班  $j$  的期望程度；

3)  $to\_visit_k$  ( $k=1,2,\dots,m$ ) 为蚂蚁  $k$  待访问城市的集合，开始时， $to\_visit_k$  有  $(n-1)$  个元素，即包括除了蚂蚁  $k$  出发城市的其他所有城市，随着时间的推进， $to\_visit_k$  中的元素不断减少，直至为空，即表示所有的城市均访问完毕。

### (3) 更新信息素

计算各个蚂蚁经过的时间长度  $L_k$  ( $k=1,2,\dots,m$ )，记录当前迭代次数的最优解（最短时间）。同时根据下面两式对各个航班间隔时间的信息素浓度进行更新。

当所有蚂蚁完成一次循环之后，各个航班之间的尾流间隔上的信息素浓度需进行更新，计算公式为：

$$\begin{cases} \tau_{ij}(t+1) = (1-\rho)\tau_{ij}(t) + \Delta\tau_{ij} \\ \Delta\tau_{ij} = \sum_{k=1}^n \Delta\tau_{ij}^k \end{cases}, 0 < \rho < 1 \quad (8)$$

其中  $\Delta\tau_{ij}^k$  表示第  $k$  只蚂蚁在航班  $i$  与航班  $j$  连接路径上的信息素浓度； $\Delta\tau_{ij}$  表示所有蚂蚁在航班  $i$  与航班  $j$  尾流间隔之间连接路径上的信息素浓度之和。

$$\tau_{ij}^k = \begin{cases} \frac{Q}{L_k}, & \text{第 } k \text{ 只蚂蚁从航班 } i \text{ 访问航班 } j \\ 0, & \text{其他} \end{cases} \quad (9)$$

其中  $Q$  为常数，表示蚂蚁循环一次所释放的信息素总量； $L_k$  表示第  $k$  只蚂蚁经过的时间长度。

### (4) 判断是否终止

若  $NC < NC_{\max}$ ，则令  $NC = NC + 1$ ，清空蚂蚁经过路径的记录表，并返回步骤 (2)，否则终止计算，输出最优解。

选取深圳机场某日 9:00–9:30 之间的延误航班为实验数据，如下表：

表 4.21 航班数据

编号	航班号	预计起飞时间	实际起飞时间	机型	最大载客人数
1	CZ3553	9:00	10:06	重型	440
2	CZ6310	9:00	16:47	中型	180
3	CZ6969	9:00	9:48	中型	220
4	CZ3663	9:05	10:08	中型	220
5	CZ3978	9:10	10:18	中型	220
6	HU7745	9:10	11:40	中型	189
7	ZH9867	9:10	10:21	中型	189
8	CZ6236	9:15	12:02	中型	189
9	HU7763	9:15	11:48	中型	189
10	DZ6253	9:20	10:52	中型	189
11	HU7107	9:20	10:40	中型	189
12	HU7163	9:20	10:09	重型	440
13	HU775Z	9:20	10:42	中型	440
14	ZH9737	9:25	11:58	中型	189

由表中可知最早恢复的寒航班起飞时间为 9:48，假设，9:45 开始延误航班的恢复。

表 4.22 航班延误造成的经济损失

编号	航班号	机型	最大载客人数	运营成本/元 每小时	延误时间/分钟	经济损失/元
1	CZ3553	重型	440	4167	66	36469.62
2	CZ6310	中型	180	2916	467	114994.1
3	CZ6969	中型	220	2916	48	13927.68
4	CZ3663	中型	220	2916	63	18280.08
5	CZ3978	中型	220	2916	68	19730.88
6	HU7745	中型	189	2916	150	38418.3
7	ZH9867	中型	189	2916	71	18184.66
8	CZ6236	中型	189	2916	167	42772.37
9	HU7763	中型	189	2916	153	39186.67
10	DZ6253	中型	189	2916	92	23563.22
11	HU7107	中型	189	2916	80	20489.76
12	HU7163	重型	440	4167	49	27075.93
13	HU775Z	中型	440	2916	82	43601.04
14	ZH9737	中型	189	2916	153	39186.67
总计						495881

延误时间均在半个小时以上，按照上文中介绍的延误成本算法进行计算，在这段时间内，按实际的情况，航班延误造成的经济成本高达近 50 万元。

在 MATLAB 中，将实验数据代入编号的蚁群算法航班延误优化程序中，对这 14 班航班进

行起飞序列排序的优化运算, 得到:

迭代情况如下:

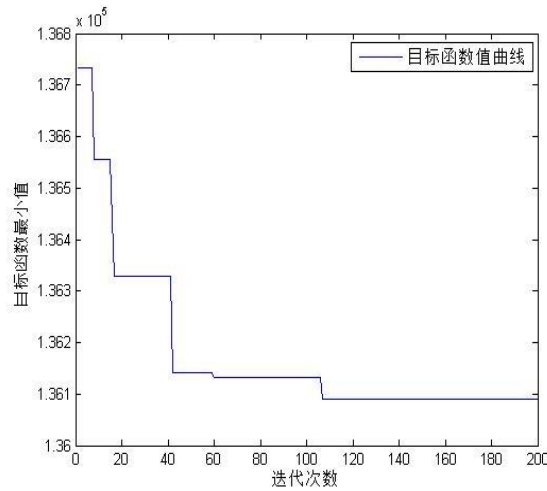


图 4.16 迭代过程

优化的航班起飞排序编号序列为:

13, 10, 1, 12, 2, 3, 6, 4, 5, 7, 8, 11, 9, 14

最小延误成本为 136090.3 元, 运算时间 7.77994 秒。

说明可在极短时间内得出恢复延误航班起飞排序的优化方案, 且若按照方案实行, 可使得航班延误的时间和所造成的经济损失都得到有效的减小, 仅对这短短半个小时内的延误航班进行恢复排序就可以将延误成本降低 30 多万元, 若将此算法投入到日常航班排序中, 并严格执行, 每年我国在延误航班上的经济损失必能大幅减少。

不可忽略的是, 实际情况中, 受其他的起落航班和其他一些因素的影响, 可能难以将机场的某跑道的某段时间都用于延误航班的恢复, 因而影响航班恢复的实际效果。若可将正常的起降航班也都加入排序会更符合实际情况, 本文受数据收集的限制, 难以做到对所有航班 (包括正常与不正常的航班) 进行排序。

为了计算使用此算法每年大约减少的延误航班造成的损失, 考虑到不同时间的航班延误数量的不同, 我们想到利用延误航班总数和平均为每架延误航班节约的成本来进行估算。根据以上仿真实验的结果, 假设使用此算法可平均为每架延误航班减少 2.5 万元 (由将降低的大约 35 万元平摊到 14 架延误航班的延误成本上而得) 的损失。

我们根据全球领先的航班数据服务商 VariFlight (飞友科技) 发布《全球机场放行准点率报告》中关于深圳宝安机场过去一年 (由 2014 年 6 月到 2015 年 5 月) 的月度航班数据, 计算得到每月的晚点航班数量, 再计算得每个月的以及过去一年的能节省的延误航班成本, 结果显示, 每年约能为深圳机场减少约 15 亿人民币的延误航空成本, 若投入到全国各个机场, 必能有效降低我国每年巨额的延误航班所带来的经济损失。

表 4.23 过去一年深圳宝安机场的延误航班经济损失估算

时间	样本量	准点	略晚点	晚点	严重晚点	总晚点率	晚点航班数量	节省延误成本/万元
2014 年 6 月	10149	60.09%	11.77%	6.93%	21.20%	39.91%	4050	10125
2014 年 7 月	11264	48.22%	12.02%	7.96%	31.79%	51.78%	5832	14580
2014 年 8 月	11676	43.12%	13.16%	8.60%	35.12%	56.88%	6641	16602.5
2014 年 9 月	11039	54.33%	13.11%	8.30%	24.27%	45.68%	5042	12605
2014 年 10 月	11757	69.49%	11.48%	6.85%	12.18%	30.51%	3587	8967.5
2014 年 11 月	11579	64.95%	13.69%	8.15%	13.21%	35.05%	4059	10147.5
2014 年 12 月	11218	71.73%	11.64%	6.18%	10.45%	28.27%	3171	7927.5
2015 年 1 月	11555	67.63%	13.12%	6.84%	12.41%	32.37%	3740	9350
2015 年 2 月	11475	54.00%	14.37%	8.31%	23.32%	46.00%	5278	13195
2015 年 3 月	12036	60.37%	13.43%	8.45%	17.75%	39.63%	4770	11925
2015 年 4 月	11498	51.41%	12.65%	8.75%	27.19%	48.59%	5587	13967.5
2015 年 5 月	11068	38%	10%	6%	45%	62.02%	6864	17160
总计								146552.5

结果显示，每年约能为深圳机场减少约 15 亿人民币的延误航空成本，若投入到全国的各个机场，必能有效降低我国每年巨额的延误 航班所带来的经济损失。

五、模型评价

5.1 对问题一：

优点：

- （1）从多角度考虑可能性，对数据处理，进行直观的对比与分析。
- （2）层次分析法和熵权法的结合运用将主观与客观结合来看，更符合实际情况

缺点：

- （1）在对机场滑行时间的对比上，由于 13 年数据较少，未能充分对比。可能会出现与猜想假设不符的情况。
- （2）层次分析法中的准则层并未完全概括所有因素，可能存在系统误差。

5.2 主成分分析：

优点：这种方法避免了在综合评分等方法中权重确定的主观性和随意性，评价结果比较符合实际情况；同时，主成份分量表现为原变量的线性组合，如果最后综合指标包括所有分量，则可以得到精确的结果，百分之百地保留原变量提供的变差信息，即使舍弃若干分量，也可以保证将 85%以上的变差信息体现在综合评分中，使评价结果真实可靠。是在实际中应用得比较广的一种方法。由于其第一主成份（因子）在所有的主成分中包含信息量最大，很多学者在研究综合评价问题时常采用第一主成分来比较不同实体间的差别。

综上所述，该方法的优点主要体现在两个方面：

- （1）权重确定的客观性；
- （2）评价结果真实可靠。

缺点：主成分的解释其含义一般多少带有点模糊性，不像原始变量的含义那么清楚、确切，这是变量降维过程中不得不付出的代价。因此，提取的主成分个数  $m$  通常应明显小于原始变量个数  $p$ （除非  $p$  本身较小），否则维数降低的“利”可能抵不过主成分含义不如原始变量清楚的“弊”

5.3 蚁群算法

优点：

- (1) 采用正反馈机制，使得搜索过程不断收敛，最终逼近最优解。
- (2) 每个个体可以通过释放信息素来改变周围的环境，且每个个体可以感知周围环境的实时变化，个体间通过环境进行间接地通讯。
- (3) 搜索过程采用分布式计算方式，多个个体同时进行并行计算，大大提高了算法的计算能力和运行效率。启发式的概率搜索方式不容易陷入局部最优，易于寻找到全局最优解。

缺点：

虽然传统的蚁群算法具有很强的全局寻优解 的能力，但也存在搜索时间过长，在执行过程中容易出现停滞现象、当问题规模较大时存在陷入局部最优的可能性等缺陷。

模型设计的缺点及改善：模型设计时，由于受数据收集的限制，未能做到对所有航班（包括正常与不正常的航班）进行排序。然而在实际情况中，受其他的起落航班和其他一些因素的影响，可能难以将机场的某跑道的某段时间都用于延误航班的恢复，因而影响航班恢复的实际效果。若可将正常的起降航班也都加入排序会更符合实际情况

## 六、结语

飞机由于速度快、跨度大等特点成为了现代人出行的主要工具之一，随着对空运需求量的增大，航班延误问题受到各界广泛关注。我国的机场延误率在国外的报道上一直居高不下，而国内的统计数据则表明并非如此。通过查阅资料、数据对比我们发现国内外对延误的定义是不同的，而我国航班延误率虽然很高，但由于数据处理问题，并不像国外所报道的那么低。通过主成分分析，我们找出了航班延误的主要原因为空管原因和机场原因，在知道原因的基础上，我们则可以针对具体原因做出具体的改进措施。而针对飞机调度问题，我们通过蚁群算法，进行模拟仿真，发现合理的调度飞机是有效减少延误情况的措施。在正视延误情况严重下，进行合理分析延误原因，通过航空公司、机场、政府的多方努力，是可以使得我国的延误情况得到改善的。现如今我们就可以看到通过扩大机场的容量、增加跑道、不限起飞等措施来降低延误率的积极举动，因此我们相信，在各界的努力下，通过科学合理的方法，找出延误主因，并就各个原因逐项击破，是可以在不久的将来大幅提升正点率的。



参考文献:

- [1]Te-Ping Chen. China Airports World' s Worst for Delays[EB/OL]. South China Morning Post, <http://www.nanzao.com/sc/search?keyword=FlightStats>, 2013-7-12.
- [2] Airline and Airport On-time Performance Report[R].2013
- [3]2013 年全国民航航班运行效率报告[R]. 2014
- [4]Flightstats 延误航班统计疑似双重标准[J/OL]. 中国民航管理干部学院 民航企业管理研究基地信息中心(新闻晨报). 2013-7-15.
- [5]王明涛. 指标综合评价中权系数的一种综合分析方法[J]. 系统工程, 1999, (02):56-61.
- [6]刘艳红, 高林, 李耀华. 基于经济损失的航班恢复模型研究[N]. 中国民航大学学, 2011, 29 (1)
- [7]李志荣, 张兆宁. 基于蚁群算法的航班着陆排序[N]. 交通运输工程与信息学报, 2006, 4(2)
- [8]吴国强. 蚁群算法在航班进离港调度的应用[D]. 天津: 中国民航大学, 2007.
- [9]李冠彬, 詹志辉, 张军. 蚁群算法优化到达航班排序和调度问题的研究[N]. 计算机工程与设计, 2009, 30(17).
- [10]张启钱, 胡明华, 施赛峰. 基于 RHC 的航班着落调度多目标优化算法[N]. 南京航空航天大学学报, 2012, 44(3).
- [11]徐兆龙, 姜丽, 罗宇骁, 徐新星. 基于蚁群算法的多跑道航班协同调度建模[N]. 武汉理工大学学报(交通科学与工程版), 2014, 38(6).
- [12]史峰, 王辉, 郁磊, 胡斐. MATLAB 智能算法 30 个案例分析[M]. 北京:北京航空航天大学出版社, 2011.
- [13]薛薇. SPSS 统计分析方法及应用[M]. 北京: 电子工业出版社, 2013.
- [14]张静. 2004 年世界前 100 家航空公司经营状况[J]. 中国民用航空. 2005. 6(11): 27—31.
- [15]2013 年全国民航航班运行效率报告[R]. 2014
- [16]李旭宏, 李玉民, 顾政华, 杨文东. 基于层次分析法和熵权法的区域物流发展竞争态势分析[N]. -东南大学学报(自然科学版) 2004(3)
- [17]杨秀云, 王军, 何建宝. 航班延误关键影响因素及影响程度识别——基于动态排队模型的分析[N]-统计与信息论坛 2014(4)

附录

AHP 一致性检验结果:

1. 我国机场延误状况    一致性比例: 0.0836; 对“我国国际产延误状况”的权重: 1.0000;  $\lambda_{\max}$ :3.0869

我国机场延误状况	流量	机场	其他	Wi
流量	1	3	0.25	0.2311
机场	0.3333	1	0.2	0.1038
其他	4	5	1	0.6651

2. 流量    一致性比例: 0.0286; 对“我国机场延误状况”的权重: 0.2311;  $\lambda_{\max}$ : 5.1283

流量	货邮吞吐量	旅客吞吐量	起降架次	日均起降架次	日高峰架次	Wi
货邮吞吐量	1	0.3333	0.2	0.2	0.1429	0.0436
旅客吞吐量	3	1	0.3333	0.3333	0.2	0.0886
起降架次	5	3	1	1	0.3333	0.2017
日均起降架次	5	3	1	1	0.3333	0.2017
日高峰架次	7	5	3	3	1	0.4641

3. 机场    一致性比例: 0.0000; 对“我国机场延误状况”的权重: 0.1038;  $\lambda_{\max}$ : 2.0000

机场	平均出港滑行时间	跑到数量	Wi
平均出港滑行时间	1	0.3333	0.25
跑到数量	3	1	0.75

4. 其他    一致性比例: 0.0000; 对“我国机场延误状况”的权重: 0.6651;  $\lambda_{\max}$ : 2.0000

机场	平均出港滑行时间	跑到数量	Wi
GDP	1	0.1111	0.1
天气	9	1	0.9

5. 货邮吞吐量 一致性比例: 0.0416; 对“我国机场延误状况”的权重: 0.0101;  $\lambda$

max: 12.7040

货 邮 吞 吐 量	PEK	CAN	PVG	SHA	CTU	SZX	KMG	XIY	CKG	HGH	TSN	DLC	Wi
PEK	1	1	0.33	4	4	2	6	9	6	6	8	9	0.17
CAN	1	1	0.25	3	3	1	5	9	5	5	7	9	0.15
PVG	3	4	1	7	7	4	8	9	8	8	8	9	0.3
SHA	0.25	0.33	0.14	1	1	0.5	1	4	1	1	2	4	0.05
CTU	0.25	0.33	0.14	1	1	0.5	1	4	1	1	2	4	0.05
SZX	0.5	1	0.25	2	2	1	0.33	6	3	3	4	6	0.09
KMG	0.17	0.2	0.13	1	1	3	1	3	1	1	2	3	0.06
XIY	0.11	0.11	0.11	0.25	0.25	0.17	0.33	1	0.33	0.33	0.5	1	0.11
CKG	0.17	0.2	0.13	1	1	0.33	1	3	1	1	2	3	0.04
HGH	0.17	0.2	0.13	1	1	0.33	1	3	1	1	2	3	0.04
TSN	0.13	0.14	0.13	0.5	0.5	0.25	0.5	2	0.5	0.5	1	2	0.02
DLC	0.11	0.11	0.11	0.25	0.25	0.17	0.33	1	0.33	0.33	0	1	0.02

6. 旅客吞吐量 一致性比例: 0.0139; 对“我国机场延误状况”的权重: :0.0205;  $\lambda$  max: 12.2362

旅 客 吞 吐 量	PEK	CAN	PVG	SHA	CTU	SZX	KMG	XIY	CKG	HGH	TSN	DLC	Wi
PEK	1	2	2	3	3	3	3	3	3	4	8	7	0.2
CAN	0.5	1	1	2	2	2	3	3	3	4	7	6	0.15
PVG	0.5	1	1	2	2	2	3	3	3	4	7	6	0.15
SHA	0.33	0.5	0.5	1	1	1	1	2	2	3	5	4	0.08
CTU	0.33	0.5	0.5	1	1	1	1	2	2	3	5	4	0.08
SZX	0.33	0.5	0.5	1	1	1	1	2	2	3	5	4	0.08
KMG	0.33	0.33	0.33	1	1	1	1	2	2	3	4	3	0.08
XIY	0.33	0.33	0.33	0.5	0.5	0.5	0.5	1	1	1	4	3	0.05
CKG	0.33	0.33	0.33	0.5	0.5	0.5	0.5	1	1	1	3	2	0.05
HGH	0.25	0.25	0.25	0.33	0.33	0.33	0.33	1	1	1	3	2	0.04
TSN	0.13	0.14	0.14	0.2	0.2	0.2	0.2	0.25	0.33	0.33	1	1	0.02
DLC	0.14	0.17	0.17	0.25	0.25	0.25	0.25	0.33	0.5	0.5	1	1	0.02

7. 起降架次      一致性比例: 0.0067; 对“我国机场延误状况”的权重: :0.0466;  $\lambda_{\max}$ : 12.1132

起降架次	PEK	CAN	PVG	SHA	CTU	SZX	KMG	XIY	CKG	HGH	TSN	DLC	Wi
PEK	1	2	2	3	3	3	3	4	4	4	6	6	0.21
CAN	0.5	1	1	2	2	2	2	3	3	3	5	5	0.14
PVG	0.5	1	1	2	2	2	2	3	3	3	5	5	0.14
SHA	0.33	0.5	0.5	1	1	1	1	1	2	2	3	3	0.08
CTU	0.33	0.5	0.5	1	1	1	1	1	2	2	3	3	0.08
SZX	0.33	0.5	0.5	1	1	1	1	1	2	2	3	3	0.08
KMG	0.33	0.5	0.5	1	1	1	1	1	2	2	3	3	0.08
XIY	0.25	0.33	0.33	1	1	1	1	1	1	1	2	2	0.06
CKG	0.25	0.33	0.33	0.5	0.5	0.5	0.5	1	1	1	2	2	0.05
HGH	0.25	0.33	0.33	0.5	0.5	0.5	0.5	1	1	1	2	2	0.05
TSN	0.17	0.2	0.2	0.33	0.33	0.33	0.33	0.5	0.5	0.5	1	1	0.03
DLC	0.17	0.2	0.2	0.33	0.33	0.33	0.33	0.5	0.5	0.5	1	1	0.03

8. 日均起降架次      一致性比例: 0.0042; 对“我国机场延误状况”的权重: :0.0466;  $\lambda_{\max}$ : 12.0714

日均起降架次	PEK	CAN	PVG	SHA	CTU	SZX	KMG	XIY	CKG	HGH	TSN	DLC	Wi
PEK	1	2	1	3	4	3	3	2	3	3	2	4	0.17
CAN	0.5	1	0.5	2	3	2	2	1	2	2	1	3	0.1
PVG	1	2	1	3	4	3	3	2	3	3	2	4	0.17
SHA	0.33	0.5	0.33	1	2	1	1	0.5	1	1	0.5	2	0.06
CTU	0.25	0.33	0.25	0.5	1	0.5	0.5	0.33	0.5	0.5	0.33	1	0.03
SZX	0.33	0.5	0.33	1	22	1	1	0.5	1	1	0.5	2	0.06
KMG	0.33	0.5	0.33	1	2	1	1	0.5	1	1	0.5	2	0.06
XIY	0.5	1	0.5	2	3	2	2	1	2	2	1	3	0.1
CKG	0.33	0.5	0.33	1	2	1	1	0.5	1	1	0.5	2	0.06
HGH	0.33	0.5	0.33	1	2	1	1	0.5	1	1	0.5	2	0.06
TSN	0.5	1	0.5	2	3	2	2	1	2	2	1	3	0.1
DLC	0.25	0.33	0.25	0.5	1	0.5	0.5	0.33	0.5	0.5	0.33	1	0.03

9. 日高峰起降架次 一致性比例: 0.0147; 对“我国机场延误状况”的权重: 0.1073;  $\lambda_{\max}$ : 12.2494

日高峰起降架次	PEK	CAN	PVG	SHA	CTU	SZX	KMG	XIY	CKG	HGH	TSN	DLC	Wi
PEK	1	2	2	4	3	3	3	4	4	4	6	6	0.22
CAN	0.5	1	1	1	2	2	2	3	3	3	4	4	0.13
PVG	0.5	1	1	1	2	2	2	3	3	3	4	4	0.13
SHA	0.25	1	1	1	0.5	0.5	0.5	1	1	1	2	2	0.06
CTU	0.33	0.5	0.5	2	1	1	1	2	2	2	3	3	0.09
SZX	0.33	0.5	0.5	2	1	1	1	2	2	2	3	3	0.09
KMG	0.33	0.5	0.5	2	1	1	1	2	2	2	3	3	0.09
XIY	0.25	0.33	0.33	1	0.5	0.5	0.5	1	1	1	2	2	0.05
CKG	0.25	0.33	0.33	1	0.5	0.5	0.5	1	1	1	2	2	0.05
HGH	0.25	0.33	0.33	1	0.5	0.5	0.5	1	1	1	2	2	0.05
TSN	0.17	0.25	0.25	0.5	0.33	0.33	0.33	0.5	0.5	0.5	1	1	0.03
DLC	0.17	0.25	0.25	0.5	0.33	0.33	0.33	0.5	0.5	0.5	1	1	0.03

10. 平均出港滑行时间 一致性比例: 0.0071; 对“我国机场延误状况”的权重: 0.0260;  $\lambda_{\max}$ : 12.1202

平均出港滑行时间	PEK	CAN	PVG	SHA	CTU	SZX	KMG	XIY	CKG	HGH	TSN	DLC	Wi
PEK	1	2	2	3	3	3	3	4	4	4	5	5	0.21
CAN	0.5	1	1	2	2	2	2	3	3	3	4	4	0.14
PVG	0.5	1	1	2	2	2	2	3	3	3	4	4	0.14
SHA	0.33	0.5	0.5	1	1	1	1	2	2	2	3	3	0.08
CTU	0.33	0.5	0.5	1	1	1	1	2	2	2	3	3	0.08
SZX	0.33	0.5	0.5	1	1	1	1	2	2	2	3	3	0.08
KMG	0.33	0.5	0.5	1	1	1	1	2	2	2	3	3	0.08
XIY	0.25	0.33	0.33	0.5	0.5	0.5	0.5	1	1	1	2	2	0.05
CKG	0.25	0.33	0.33	0.5	0.5	0.5	0.5	1	1	1	2	2	0.05
HGH	0.25	0.33	0.33	0.5	0.5	0.5	0.5	1	1	1	2	2	0.05
TSN	0.2	0.25	0.25	0.33	0.33	0.33	0.33	0.5	0.5	0.5	1	1	0.03
DLC	0.2	0.25	0.25	0.33	0.33	0.33	0.33	0.5	0.5	0.5	1	1	0.03

11. 跑道 一致性比例: 0.0000; 对“我国机场延误状况”的权重: 0.0779;  $\lambda_{\max}$ : 12.0000

跑道	PEK	CAN	PVG	SHA	CTU	SZX	KMG	XIY	CKG	HGH	TSN	DLC	Wi
PEK	1	2	1	2	2	2	2	2	2	2	2	2	0.14
CAN	0.5	1	0.5	1	1	1	1	1	1	1	1	1	0.07
PVG	1	2	1	2	2	2	2	2	2	2	2	2	0.14
SHA	0.5	1	0.5	1	1	1	1	1	1	1	1	1	0.07
CTU	0.5	1	0.5	1	1	1	1	1	1	1	1	1	0.07
SZX	0.5	1	0.5	1	1	1	1	1	1	1	1	1	0.07
KMG	0.5	1	0.5	1	1	1	1	1	1	1	1	1	0.07
XIY	0.5	1	0.5	1	1	1	1	1	1	1	1	1	0.07
CKG	0.5	1	0.5	1	1	1	1	1	1	1	1	1	0.07
HGH	0.5	1	0.5	1	1	1	1	1	1	1	1	1	0.07
TSN	0.5	1	0.5	1	1	1	1	1	1	1	1	1	0.07
DLC	0.5	1	0.5	1	1	1	1	1	1	1	1	1	0.07

12. GDP 一致性比例: 0.0090; 对“我国机场延误状况”的权重: 0.0665;  $\lambda_{\max}$ : 12.1533

GDP	PEK	CAN	PVG	SHA	CTU	SZX	KMG	XIY	CKG	HGH	TSN	DLC	Wi
PEK	1	2	0.5	0.5	3	2	4	4	2	3	2	3	0.12
CAN	0.5	1	0.33	0.33	2	1	3	3	1	2	1	2	0.08
PVG	2	3	1	1	4	3	5	5	3	4	3	4	0.19
SHA	2	3	1	1	4	3	5	5	4	5	3	4	0.2
CTU	0.33	0.5	0.25	0.25	1	0.5	2	2	0.5	1	0.5	1	0.04
SZX	0.5	1	0.33	0.33	2	1	3	3	1	2	1	2	0.08
KMG	0.25	0.33	0.2	0.2	0.5	0.33	1	1	0.33	0.5	0.33	0.5	0.03
XIY	0.25	0.33	0.2	0.2	0.5	0.33	1	1	0.33	0.5	0.33	0.5	0.03
CKG	0.5	1	0.33	0.33	2	1	3	3	1	2	1	2	0.07
HGH	0.33	0.5	0.25	0.25	1	0.5	2	2	0.5	1	0.5	1	0.04
TSN	0.5	1	0.33	0.33	2	1	3	3	1	2	1	2	0.08
DLC	0.33	0.5	0.25	0.25	1	0.5	2	2	0.5	1	0.5	1	0.04

