



第三军医大学

2015 “深圳杯” 数学建模夏令营论文

题目：航班的延误成因分析及其优化排列模型

题目序号：D18

指导老师：罗万春

小组成员：李翔 凌萍 陈琳

学校：第三军医大学

二〇一五年八月

航班的延误成因分析及其优化排列模型

摘要

本文研究的是航班延误问题，重点讨论航班准点标准的合理性，航班延误成因以及解决方案等内容。

对于问题一，首先搜集了全球 15 家机场航班的数据，并且在处理数据时将湿租业务与代码共享业务范围内的同一航班视为多家公司的多次航班。然后分别以航班 15min（美国标准）、30min（中国标准）、45min 内到达作为准点标准，以延误率作为衡量延误严重程度的模型，得到不同标准下延误率最高的三家机场均为上海虹桥国际机场，伦敦希思罗国际机场和首尔金浦国际机场。其次，对三种标准下的 15 家机场的排名进行 K-W 秩和检验及中美两国标准下的机场排名相关性分析，发现三种标准下的排名具有显著一致性，并且中美两国标准下的排名还具有显著线性相关性。最后，我们通过对中美两国航班延误成因的种类划分，各成因所占比例及其逐年变化进行比较分析，发现两国在成因划分上有较大相似度。并且航空公司自身原因作为中国航班延误的主要成因，其所占比例与美国相差不大，而各项成因所占比例也随时间改变保持稳定。由此可得出结论：使用美国标准对我国航班延误情况进行统计是合理的。因此可判定题目所给结论正确。

对于问题二，搜集了 2006-2013 年中国航班延误成因及其所占比例，分别采用熵权法与主成分分析法来分析航班延误的主要成因。求解结果为：航空公司自身原因与流量控制在航班延误中所占权重（分别为 0.5791 和 0.2557）较大，对航班延误贡献率（分别为 62.82%，30.00%）较高。而天气与其他成因在航班延误中所占权重（分别为 0.1676，0.0024）较小，对航班延误贡献率（分别为 7.18%，0.00%）较低。由此可得出结论：我国航班延误的主要成因因为航空公司自身原因与不科学的流量控制。

对于问题三，由于我国航班延误的主要成因是航空公司自身原因与不科学的流量控制，因此着重从航班调度的角度对我国航班延误情况提出改进措施。以机场单跑道为例进行讨论，建立了以机场现有航班着陆或起飞所消耗总时间最少为目标函数，保证各航班能连续降落及安全间隔时间等为约束条件的非线性优化模型，并使用深圳宝安国际机场于 2015 年 5 月 12 日 14:00-15:00 的飞机着陆数据，利用模拟退火遗传算法求解出着陆总时间最小的飞机排列顺序为 9、1、8、2、7、10、4、3、5、6 以及各飞机着陆的具体时间，并分析了该结论的可靠性。

本文建立的基于指标权重分析和模拟退火遗传算法的航班优化排列模型，能较好地解决航班延误问题，在路径规划、系统设计等方面有推广应用价值。

关键词：航班延误 权重分析 模拟退火遗传算法 最优策略

1 问题重述

1.1 问题背景

香港南华早报网根据 flightstats.com 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。

1.2 数据集[1-4]

以下数据均为在官方媒体上所搜集到的相关航班延误情况、延误原因的数据，具体内容见文内或压缩包附件。

- (1) flightstats.com 网站 2015 年 5 月 2 日上海虹桥机场等 15 家机场航班数据；
- (2) 权威新闻媒体公布 2006-2008 年国内航班延误原因数据；
- (3) 2009-2013 民航行业发展统计公报；
- (4) 美国运输部公布的航空相关数据；
- (5) 深圳宝安机场 2015 年 5 月 12 日到港航班相关数据。

1.3 提出问题

根据上述问题背景即数据，题目要求我们建立数学模型讨论下列问题。

- (1) 题目给出的结论是否正确？
- (2) 我国航班延误的主要原因是什么？
- (3) 有什么改进措施？

2 模型假设

- (1) 将航空公司间湿租业务与共享代码业务下同一航班视为多家公司的多次航班。
- (2) 飞机的最早（晚）到达时间采用之前一周中同一航班飞机的最早（晚）到达时间。
- (3) 假设飞机离港或到港调度只受航空公司和国家流量控制原因的影响。
- (4) 假设各航空公司官网和相关媒体以及相关统计网站公布的数据真实可靠。

注：共享代码业务与湿租业务定义在文中注解

3 符号说明

u ：所挑选的机场序号 ($u = 1, 2, \dots, 15$)

α_{uv} ：在 v 标准下第 u 家机场的延误率 ($u = 1, 2, \dots, 15; v = 1, 2, 3$)

X_{uw} ：在第 w 延误等级内第 u 家机场航班到达数量 ($u = 1, 2, \dots, 15; w = 1, 2, 3, 4$)

n : 流量控制时终端区即将到达的飞机总数

i, j : 到达飞机序号, (i 飞机紧跟飞机 j 着陆)

T_{ij} : 飞机 i, j 之间必须保持最小时间间隔

E_i : 飞机 i 能够降落的最早时间

L_i : 飞机 i 能够降落的最晚时间

$[E_i, L_i]$: 飞机 i 可能降落的时间窗

ETA_i : 飞机 i 预计到达时间

STA_i : 飞机 i 调度到达时间

t_j : 飞机 j 的理想调度时间

x_{ij} : 决策变量, 当 $x_{ij} = 1$ 时, 表示飞机 j 紧跟飞机 i , 当 $x_{ij} = 0$ 时, 表示飞机 i 紧跟飞机 j 。 $(x_{ij} \in \{0, 1\})$

注: 其余符号在文中使用时说明。

4 问题分析

4.1 问题一

题目给出一则新闻报道, 让我们通过自行查找的数据论证“国际航班延误最严重的 10 个机场中, 中国占了 7 个”的说法是否正确。

首先, 根据我们查阅的资料显示: 中美对于延误标准存在差异, 并且由于中美两国政策, 气候, 民航空中利用率等条件均不同, 故以美国的延误标准衡量中国航班是否延误是不妥的, 因此报道中所给结论的正确性有待考量。

其次, 为验证美国标准衡量中国航班延误的合理性, 我们在 flightstats 网站随机搜集了全球 15 家机场于 2015 年 5 月 2 日各航班延误情况的相关数据[1], 利用国际上通用的对于延误等级的划分标准对当日航班延误率进行统计和排名, 通过比较不同标准下的排名即可得到该方法是否合理。

最后, 我们列出了中美两国在延误原因划分方面的相同点与不同点, 并通过观察 2011-2013 各航班延误原因及其所占比例, 以及随着时间增长各航班延误原因所占比例的变化趋势, 进一步分析判断以美国标准衡量中国航班延误情况的方法是否正确。

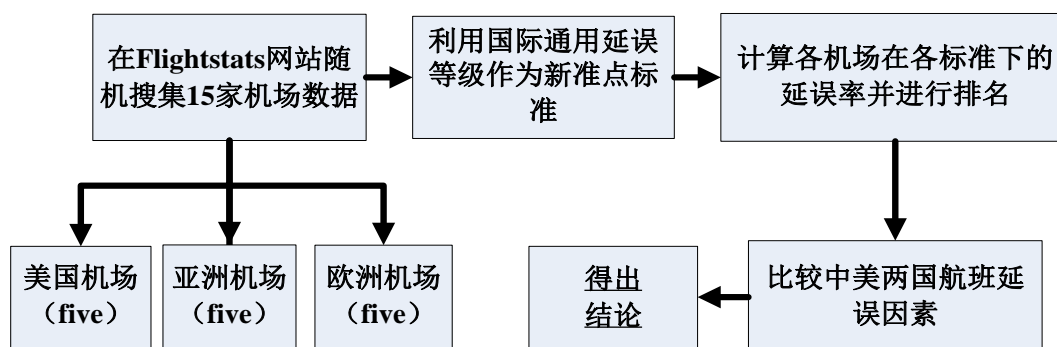


图1 思路流程图

4.2 问题二

问题二要求我们对国内航班延误的情况进行原因分析。

首先，由权威媒体公布的 2006-2008 年国内航班延误原因及其所占比例[3]，和 2009-2013 民航行业发展统计公报显示的数据[3]，可得到航班延误的原因有四点：航空公司自身原因，流量控制，天气原因以及其他。

其次，为观察各因素所占比例是否有差异性，采用单因素方差分析进行检验。但是仅根据航班延误的百分比数据和差异性结果并不能确定所有航班延误的主要原因。所以我们采用计算不同原因权重的方式找出主要原因。而能实现该过程的方法有熵权法，主成分分析法等。由于在本问题中我们搜集到的数据有指标数量少，数据量较少的特点，主要采用熵权法与主成分分析法共同分析的方式来找到影响航班延误的主要原因。

最后，根据熵权法计算出的各指标权重，以及主成分分析法所计算出各原因对航班延误的贡献率和累计贡献率可得到几个关键因素。

4.3 问题三

问题三要求我们针对问题二找出的导致我国航班延误的主要原因提出改善航班延误的方法。

首先，根据问题二可知，我国航班延误主要是由于航空公司自身原因和流量控制。这两大原因反映的是我国航空公司及相关部门对离港或到港航班调度不均进而导致航班出发或到达延误的问题。所以，应从优化航班调度的方面来考虑减少我国航班延误的发生。

其次，基于以上考虑，我们从最基本的单跑道飞机着陆的排序问题入手。将飞机在保证安全的基础上尽可能一架接连一架进行着陆或起飞时，使所有飞机着陆或起飞所花总时间最少进而使延误情况减少作为最优控制策略。根据该思路我们列出目标函数与约束条件，将该问题转变成飞机排序的优化问题。

最后，为更好弥补传统遗传算法局部搜索能力不强的缺点，并解决深圳宝安机场延误率高的问题。我们采取了模拟退火遗传算法，使用深圳宝安机场于 2015 年 5 月 12 日的数据来求得最优解。

5 模型建立与求解

5.1 问题一的模型建立与求解

5.1.1 重要航空相关术语解释

- (1) 湿租业务：航空运输企业将配备有机组人员的飞机承租给他人使用一定期限，承租期内听候承租方调遣，不论是否经营，均按一定标准向承租方收取租赁费，发生的固定费用均为承租方负担的业务。
- (2) 共享代码业务：是指一家航空公司的航班号（即代码）可以用在另一家航空公司的航班上。即旅客在全程旅行中有一段航程或全程航程是乘坐出票航空公司航班号但非出票航空公司承运的航班的。
- (3) 航班准点：中国标准是一个航班在计划起飞时间后 **30** 分钟内完成起飞（机轮离地）即为准点。美国等国家对于航班准点的标准则为，实际起飞时间在计划起飞时间之后的 **15** 分钟内。

5.1.2 航空公司延误数据整理

中美两国由于准点标准不同导致在各标准下计算得到的航班延误率也可能有很大出入。据新闻报道显示，美国 flightstats.com 网站利用美国标准对世界各国机场航班延误情况进行统计与排名，显然会令一些准点标准不同的国家排名靠前或靠后，不能客观反映出世界各国机场航班延误情况。

为验证该说法正确性，基于假设 1，在 flightstats.com 网站搜集了全球 15 家较有代表性的机场（美、亚、欧洲各 5 家）于 2015 年 5 月 2 日各航班落地时间数据[1]，并按照国际延误等级标准进行了不同标准下航班延误数量的统计。得到统计结果如表 1、表 2 所示。

表 1 全球挑选的 15 家机场代码及全称信息

序号 (num)	机场代码	机场全称
1	美-(MDW)	Chicago Midway International Airport
2	美-(DFW)	Dallas/Fort Worth International Airport
3	美-(ORD)	O'Hare International Airport
4	美-(ATL)	Hartsfield-Jackson Atlanta International Airport
5	美-(EWR)	Newark Liberty International Airport
6	亚-(GMP)	Gimpo -International Airport
7	亚-(ROK)	Rockhampton Airport
8	亚-(SHA)	上海虹桥国际机场
9	亚-(PEK)	首都国际机场
10	亚-(CTU)	成都双流国际机场
11	欧-(CDG)	Charles de Gaulle Airport
12	欧-(IST)	Istanbul ataturk airport
13	欧-(LHR)	London Heathrow Airport
14	欧-(AMS)	Amsterdam Airport Schiphol
15	欧-(DEN)	Denver International Airport

表 2 各延误等级下 15 家机场航班延误数量统计结果

Airport	on-time	Late (15-29min)	very late (30-44min)	Excessive (45min+)	unknown
美 -(MDW)	3781	64	39	134	6
美-(DFW)	2478	130	77	113	5
美-(ORD)	3520	59	39	123	6
美-(ATL)	4013	109	41	162	2
美-(EWR)	1401	49	101	103	0
亚-(GMP)	119	15	10	89	4
亚-(ROK)	23	0	0	0	0
亚-(SHA)	217	71	46	184	22
亚-(PEK)	1307	58	18	53	0
亚-(CTU)	433	49	34	63	19
欧-(CDG)	1567	207	51	97	30
欧-(IST)	590	115	66	59	23
欧-(LHR)	886	320	276	1255	3
欧-(AMS)	1364	245	121	139	124
欧-(DEN)	1212	114	59	101	2

5.1.2 不同标准下的航班延误率

根据上表结果，我们可根据国际延误等级划分标准设立 3 个准点标准等级，即 15min、30min、45min。若航班落地的实际时间与计划实际时间之差在 15min、30min 或 45min 内，则意味着在这几个标准下航班准点到达。

因此，我们可根据如下公式计算各机场在不同标准下的航班延误率：

$$\alpha_{uv} = \frac{\sum_{w=v+1}^4 X_{uw}}{\sum_{w=1}^4 X_{uw}} \quad (u=1,2,\dots,15; v=1,2,3; w=1,2,3,4) \quad (1)$$

其中，

α_{uv} ：在 v 标准下第 u 家机场的延误率；

X_{uw} ：在第 w 延误等级内第 u 家机场航班到达数量；

w ：延误等级，表示实际到达时间与预计时间之差（15min（ $w=1$ ），15~29min（ $w=2$ ），30~44min（ $w=3$ ），> 45min（ $w=4$ ））；

v ：准点标准等级，即航班在 15 分钟（ $v=1$ ），30min（ $v=2$ ），45min（ $v=3$ ）内到达即为准点；

u ：所挑选的机场序号。

经 Matlab2009a 编程（程序见附录）并进行运算，得到 15 家机场在各准点标准等级下的延误率，并将延误率由高到低进行排名，整理后如表 3 所示。

表 3 各准点标准等级下 15 家机场延误率及排名

机场	15min		30min		45min	
	排名	延误率	排名	延误率	排名	延误率
美-(MDW)	14	0.0589	14	0.043	13	0.0333
美-(DFW)	10	0.1142	10	0.0678	10	0.0403
美-(ORD)	13	0.059	13	0.0432	14	0.0328
美-(ATL)	12	0.0721	12	0.0469	11	0.0374
美-(EWR)	9	0.153	7	0.1233	8	0.0623
亚-(GMP)	3	0.481	3	0.4177	2	0.3755
亚-(ROK)	15	0	15	0	15	0
亚-(SHA)	2	0.5574	2	0.4259	3	0.3407
亚-(PEK)	11	0.0898	11	0.0494	12	0.0369
亚-(CTU)	6	0.2441	4	0.1622	4	0.1054
欧-(CDG)	8	0.1819	9	0.0758	9	0.0497
欧-(IST)	4	0.2814	5	0.1465	6	0.0692
欧-(LHR)	1	0.6755	1	0.5588	1	0.458
欧-(AMS)	5	0.2534	6	0.1305	5	0.0697
欧-(DEN)	7	0.1841	8	0.1075	7	0.0679

为观察各延误标准下航班延误率是否有差异性，我们使用 SPSS19.0 对三种标准下各机场延误排名进行多个独立样本秩和检验计算，检验结果见表 4。

表 4 不同标准下延误率的 K-W 秩和检验

	χ^2	P
结果	0.1003	0.951(>0.05)

通过表 3 和表 4 结果，我们得出：

（1）以不同颜色标出各准点标准等级下延误率前 3 名机场，这些机场无论在哪个准点标准等级下，延误率均很高；

（2）在美国准点标准（15min）与中国准点标准（30min）以及新建立的 45min 标准下，各机场排名变化不大，变化最大的也仅相差两名。

为进一步分析两种排名结果的一致性，我们利用 MATLAB2009a 软件编程，对 15min 标准下的航班延误率与 30min 标准下的航班延误率进行相关分析，得到结果如图 2 所示。

以15分钟和30分钟为延误标准的延误排名相关性分析图

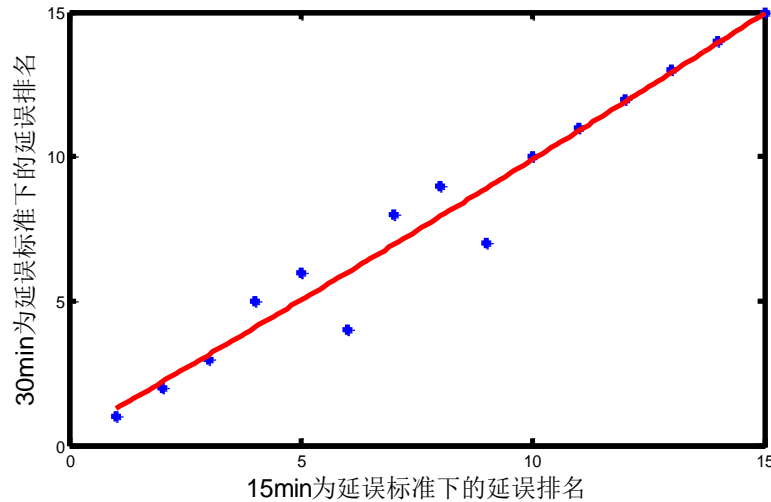


图 2 以 15min 和 30min 为标准下的延误率相关性分析

经计算，两标准下的 Spearman 等级相关的结果是： $r=0.979$ ， $P=2.65 \times 10^{-10}$ ，说明两种延误标准下的排名结果呈显著相关。

结论：使用美国定义的航班准点标准对于中国或其余国家航班是适用的。

5.1.3 中美两国航班延误原因比较

5.1.3.1 中美两国航班延误原因初步分析

根据 2011-2013 民航行业发展统计公报以及美国运输部公布的航空相关数据，我们整理得到如下信息。

(1) 中国对于航班延误原因的解释有：航空公司自身原因、流量控制、天气原因和其他。这几个原因下面还细分了诸多小原因：天气、军事活动、空管、机场、联检、油料、离岗系统、旅客及公共安全等。美国对于航班延误原因的解释有：航空公司、天气、国家航空系统、飞机晚到等。在原因划分方面，两国大致相同，因此评判两国航班延误的标准可统一。

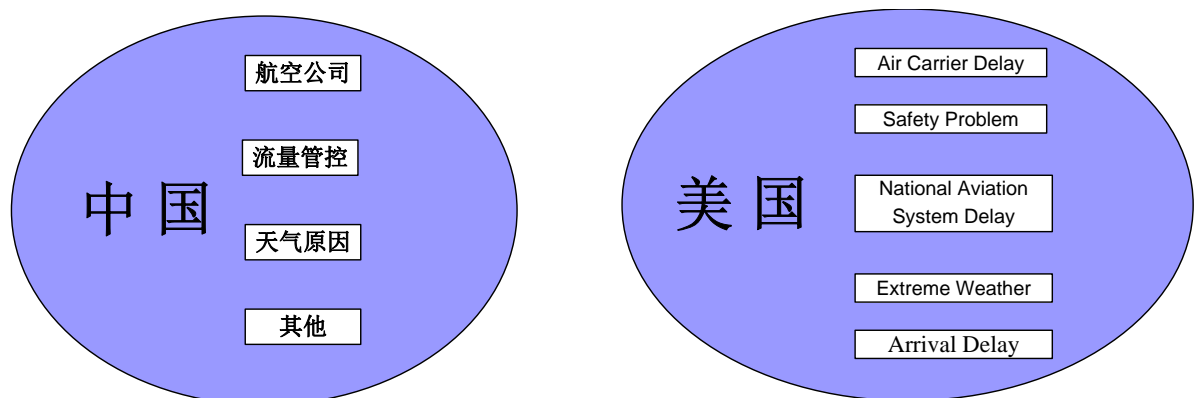


图 3 中美两国航班延误因素示意图

(2)不同延误原因对中美航班延误作用所占比例也不相同,根据所搜集资料,可作图进行对比。我们整理了 2010-2013 中美两国航班延误各原因比例(如表 5 所示)并绘制出饼状图(如图 4 所示):

表 5 中美两国 2011-2013 航班延误各原因比例						
延误原因	2011		2012		2013	
	美国	中国	美国	中国	美国	中国
航空公司原因	30.10%	37.10%	31.90%	38.50%	29.40%	37.40%
流量控制	24.80%	27.50%	22.50%	25%	24.20%	27.60%
天气原因	4.10%	20.00%	4.00%	21.60%	4.10%	21.80%
其他	40.90%	15.40%	41.50%	14.90%	42.20%	13.20%

注: 该表中, 将中国航班延误的流量控制原因归入 National Aviation System Delay (国家航空系统原因)。

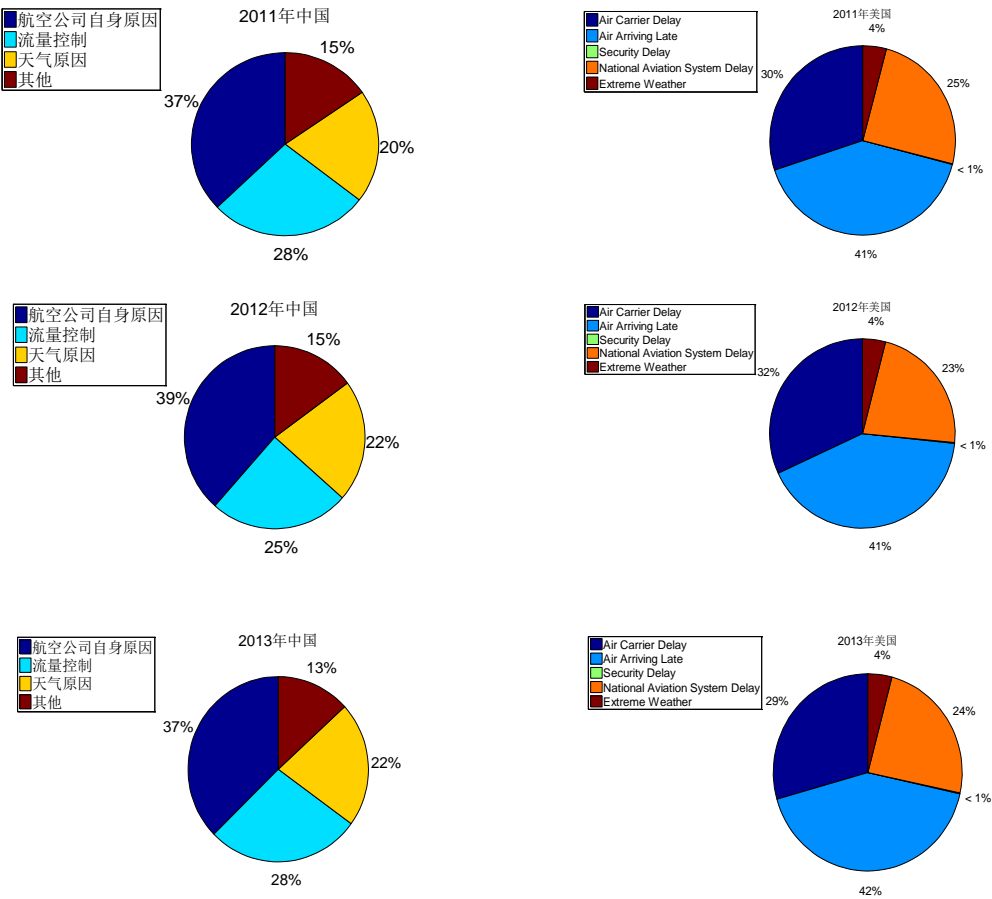


图 4 中美两国 2011-2013 航班延误各原因比例比较

由表 5 和图 4 可以看出:

- (1) 导致中美两国航班延误的原因中，两国的航空公司自身原因与国家航空系统原因所占比例相差不大；
- (2) 导致中美两国航班延误的原因中，在天气以及其他方面原因相差略大；
- (3) 中国航空公司自身原因导致的航班延误所占比例相比其他原因较大。而原因对于中国航班延误可能为主导原因，并且在这方面中美两国差异并非很大。

由此进一步证明：美国标准对中国航班延误情况进行衡量是合理的。

5.1.3.2 中美两国航班延误原因深入分析

通过搜集了 2003-2014 年美国航班延误原因及其所占比例数据[2]，以及 2006-2013 年中国航班延误原因及其所占比例数据[3]，可作出折线图以观察其趋势（如图 5、6 所示）：

2006年至2013年机场不同原因造成中国机场航班延误比例示意图

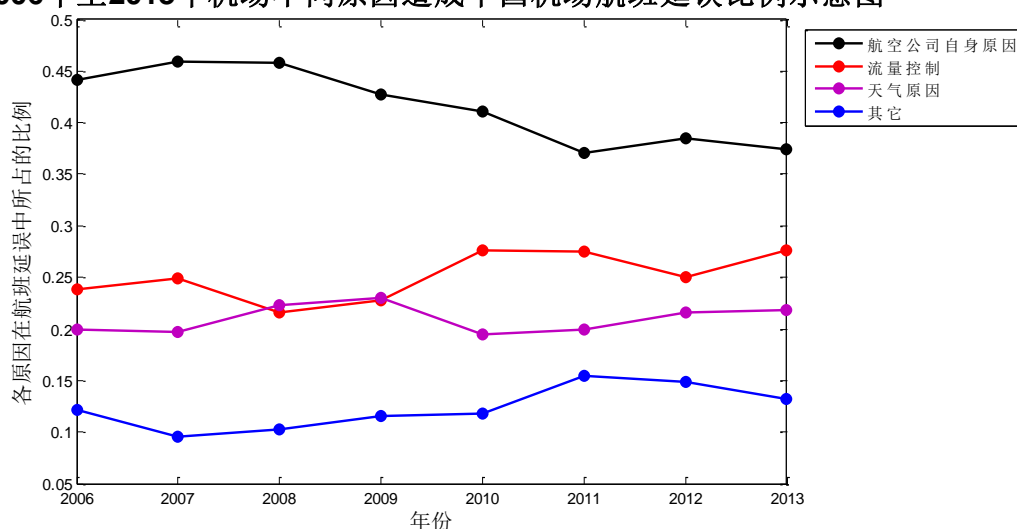


图 5 中国航班 2006-2013 年延误原因及其所占比例

2003年至2014年不同原因造成美国机场航班延误比例示意图

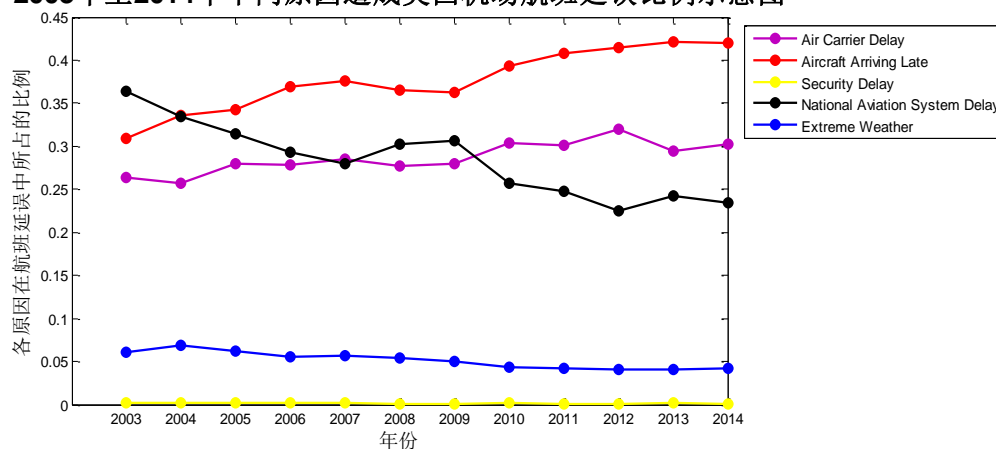


图 6 美国航班 2003-2014 年延误原因及其所占比例

为便于比较各因素的变化情况，我们计算了导致航班延误的各原因所占比例

的极差。

表 6 中国航班延误各种原因所占比例极差结果

原因	最大比例	最小比例	极差
航空公司自身原因	45.90%	37.10%	8.80%
流量控制	27.60%	21.64%	5.96%
天气原因	23.00%	19.50%	3.50%
其它	15.40%	9.50%	5.90%

表 7 美国航班延误各种原因所占比例极差结果

Reasons	Max	Min	max-min
Air Carrier Delay	31.90%	25.80%	6.20%
Others	42.30%	31.10%	8.70%
National Aviation System Delay	36.50%	22.50%	13.90%
Extreme Weather	6.90%	4.00%	2.90%

注：安全原因与航班到达晚点原因归于其他原因。

由上列图表可以看出：

(1) 随着时间增长，天气原因（最大相差 3.5%）和其他原因（最大相差 5.9%）对于中国航班延误所占比例呈较为稳定的趋势。而航空公司自身原因所占比例（最大相差 8.80%）有所下降，流量控制（最大相差 5.96%）所占比例上升，可能与近几年国家政策与机场采取相关措施有关。

(2) 随着时间增长，天气原因（最大相差 2.9%）与航空公司原因（6.2%）对于美国航班延误所占比例呈较为稳定的趋势，而流量控制（最大相差 10.9%）所占比例有较明显的下降趋势，其他原因（包含安全原因与航班到达晚点原因，最大相差 8.7%）所占比例有较为明显的上升趋势。

(3) 随着时间增长，各原因对于两国航班延误所占比例并无较大改变，特别是在天气方面。因此，我们可忽略各原因所占比例随时间变化对评判中美两国航班延误标准的影响。

5.1.4 得出结论

通过对 15 家机场各准点标准等级下的延误率进行计算和排名，并且对中美两国航班延误的原因划分以及各原因所占比例的讨论，结合表 3、4，图 1-3，可以得出如下结论：

- (1) 可以使用美国标准衡量中国或其他国家航班延误情况，美国 flightstat.com 网站公布的排名具有一定可信度；
- (2) 中国媒体和相关专家对于中国航班延误率排名较高是由于“双重标准”的解释不正确；
- (3) 香港南华早报所报道的“全球 10 个延误最严重机场中，中国占了 7 个”的情况可基本判断正确。

5.2 问题二的模型建立与求解

5.2.1 数据的搜集与整理

根据所搜集的 2009-2013 民航行业发展统计公报[3], 以及权威新闻媒体公布 2006-2008 年国内航班延误原因数据, 我们整理后得到我国 2006-2013 航班延误原因及其所占比例 (如表 8 所示):

表 8 中国 2006-2013 航班延误原因及比例

年份	航空公司自身原因	流量控制	天气原因	其它
2006	44.20%	23.80%	19.90%	12.10%
2007	45.90%	24.90%	19.70%	9.50%
2008	45.82%	21.64%	22.31%	10.23%
2009	42.72%	22.76%	23.00%	11.52%
2010	41.10%	27.60%	19.50%	11.80%
2011	37.10%	27.50%	20.00%	15.40%
2012	38.50%	25.00%	21.60%	14.90%
2013	37.40%	27.60%	21.80%	13.20%

5.2.2 各原因所占比例差异性检验

根据表 5 我们可以初步得到航空公司原因所占比例较大, 但是否各因素间具有显著差异性并不能得知。由于表 5 中各指标数据均为构成比数据, 所以我们选用卡方检验来检验航班延误原因所占比例之间是否具有显著差异。使用 SPSS19.0 运算后, 得到如下结果。

表 9 不同因素航班延误方差分析

	χ^2	P
结果	28.4143	2.97×10^{-6}

根据表 9 可以初步得到结论, 我国 2006-2013 航班延误原因所占比例之间具有显著差异性, 且航空公司原因占主导原因。但其余原因是否也在航班延误中起主要作用通过该表格无法得出结论。由于在本问题中我们搜集到的数据有指标数量较少, 数据量较少的特点, 我们将采用熵权法与主成分分析法共同分析的方式来找到影响航班延误的其他主要原因。

5.2.3 熵权法确定航班延误主要因素

5.2.3.1 熵权法背景

熵权法是一种客观赋权方法, 在具体使用过程中, 熵权法根据各指标的变异程度, 利用信息熵计算出各指标的熵权, 再通过熵权对各指标的权重进行修正, 从而得出较为客观的指标权重。

5.2.3.2 熵权法具体步骤[5]

1. 构建判断矩阵, 再对其进行标准化处理。

判断矩阵:

$$R = (r_{ij})_{m \times n} (i = 1, 2, \dots, n; j = 1, 2, \dots, m) \quad (2)$$

其中： m 表示评价的对象数， n 表示评价的指标数。

标准化：

$$b_{ij} = \frac{r_{ij} - r_{\min}}{r_{\max} - r_{\min}} \quad (3)$$

其中， r_{\max} 、 r_{\min} 分别表示同一评价指标下不同对象中最满意或最不满意者。

2. 根据熵的定义，确定这 4 个评价指标的熵值。

$$H_j = -\frac{1}{\ln n} \sum_{i=1}^n f_{ij} \ln f_{ij} \quad (4)$$

式中： $f_{ij} = b_{ij} / \sum_{i=1}^n b_{ij}, i = 1, 2, \dots, n; j = 1, 2, \dots, m; 0 \leq H_j \leq 1$ 。显然，当 $f_{ij} = 0$ 时 $\ln f_{ij}$

无意义，因此对 f_{ij} 加以修正，并对其归一化：

$$f_{ij} = \frac{1 + b_{ij}}{\sum_{i=1}^n (1 + b_{ij})} \quad (5)$$

3. 利用熵值计算评价指标的熵权：

$$W = (w_j)_{1 \times m} \quad (6)$$

$$w_j = \frac{1 - H_j}{n - \sum_{j=1}^m H_j} \quad (7)$$

5.2.3 主成分分析法确定航班延误主要因素

5.2.3.1 主成分分析法原理

在实际课题中，为了全面分析问题，往往提出很多与此有关的变量(或因素)，因为每个变量都在不同程度上反映这个课题的某些信息。而主成分分析法可将多个变量通过线性变换以选出较少个数重要变量。这种多元统计分析方法。又称主分量分析。

5.2.2.2 主成分分析法具体步骤[6]

(1) 计算协方差矩阵

计算样品数据的协方差矩阵：

$$\sum = (s_u)_{p \times p} \quad (8)$$

其中,

$$s_{ij} = \frac{1}{n-1} \sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j) \quad i, j=1,2, \dots, p \quad (9)$$

(2) 求出 Σ 的特征值 λ_i 及相应的正交化单位特征向量 a_i

Σ 的前 m 个较大的特征值 $\lambda_1 \geq \lambda_2 \geq \dots \lambda_m > 0$, 就是前 m 个主成分对应的方差, λ_i 对应的单位特征向量 a_i 就是主成分 F_i 的关于原变量的系数, 则原变量的第 i 个主成分 F_i 为:

$$F_i = a_i X \quad (10)$$

主成分的方差 (信息) 贡献率用来反映信息量的大小, α_i 为:

$$\alpha_i = \lambda_i / \sum_{i=1}^m \lambda_i \quad (11)$$

(3) 选择主成分

最终要选择几个主成分, 即 F_1, F_2, \dots, F_m 中 m 的确定是通过方差 (信息) 累计贡献率 $G(m)$ 来确定

$$G(m) = \sum_{i=1}^m \lambda_i / \sum_{k=1}^p \lambda_k \quad (12)$$

当累积贡献率大于 85% 时, 就认为能足够反映原来变量的信息了, 对应的 m 就是抽取的前 m 个主成分。

5.2.4 模型求解

根据以上提到的熵权法和主成分分析法步骤进行计算, 运用 MATLAB2009a, 得到各个指标的最终权重和对航班延误的贡献率, 见表 10、表 11。

表 10 各原因导致我国航班延误所占权重 (熵权法)

延误原因	权重
航空公司自身原因	0.5791
流量控制	0.2557
天气原因	0.1676
其它	0.0024

表 11 各原因对我国航班延误的贡献率（主成分分析法）

原因	贡献率
航空公司自身原因	62.82%
流量控制	30.00%
天气原因	7.18%
其它	0.00%

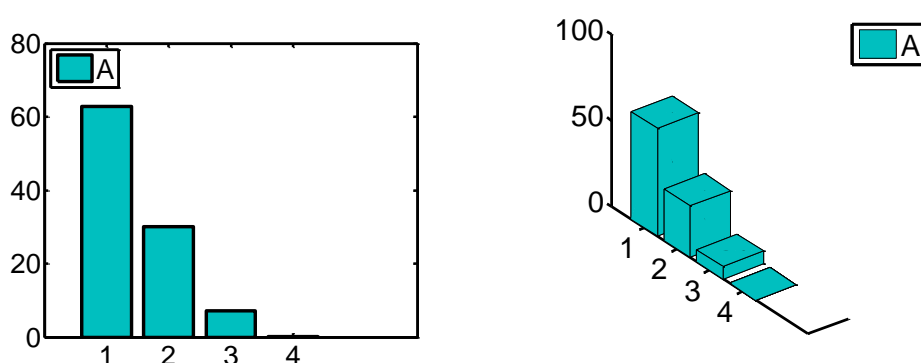


图 7 主成分分析法得到的各原因对我国航班延误的贡献率

由以上图表可以看出，航空公司自身原因与流量控制在导致航班延误的情形中所占权重（分别为 0.5791, 0.2557）最大，并且这两个原因对我国航班延误的累加贡献率已超 90%。而天气情况以及其他原因对我国航班延误影响不大。

因此，可得出结论：影响我国航班延误的主要原因为航空公司自身原因与流量控制。

5.3 问题三的模型建立与求解

5.3.1 问题背景

我国航班延误的主要原因为航空公司自身原因与流量控制原因，而这两个原因均与机场调度密切相关。因此，讨论机场如何调度才能最大程度避免我国航班出发或着陆延误是解决本题的关键所在。

5.3.2 数据的搜集与整理

本题以深圳宝安机场为例，将使用深圳宝安机场于 2015 年 5 月 12 日的数据 [4]，拟解决其延误率高的问题。

表 12 深圳宝安机场 2015 年 5 月 12 日部分航班到达数据

编号	航班	机型	机型代码	所属 型别	预达时间	实际到达时间	最早到达时间	最晚到达时间
1	南方 CZ6795	A321	1	重	13:50:00	13:54:00	13:50:00	14:11:00
2	南方 CZ3280	A321	1	重	13:55:00	14:01:00	13:52:00	14:54:00
3	国航 CA1383	B738	2	大	14:20:00	14:06:00	14:11:00	17:10:00
4	深圳 ZH9824	B738	2	大	14:15:00	14:17:00	14:15:00	14:54:00
5	东方 MU5341	A320	2	大	14:15:00	14:20:00	14:16:00	14:31:00
6	海南 HU7108	B738	2	大	14:25:00	14:26:00	14:22:00	14:57:00
7	南方 CZ3466	A321	1	重	14:30:00	14:29:00	14:15:00	15:17:00
8	深圳 ZH9507	B738	2	大	14:35:00	14:35:00	14:31:00	15:25:00
9	南方 CZ3554	A333	1	重	14:45:00	14:48:00	14:45:00	15:45:00
10	海南 HU7330	B738	2	大	14:50:00	14:56:00	14:48:00	15:14:00

注：为便于计算，在用 MATLAB 计算代入数据时，我们将 13:50:00 分视为 1350 分。

本问旨在利用相关数据对未到达机场的航班进行排序。但由于官方公布的数据量有限，且均为已到达机场的飞机数据，我们只能在已有数据的基础上对一些数据的含义赋予了新的合理定义以适应模型求解。

如上表所示，我们将实际到达时间视作航空公司调度到达时间，即不考虑飞机间安全间隔时各飞机调度后的时间。最早到达时间与最晚到达时间均为这十架飞机前一周中从相同出发点到相同目的地所花费的最少时间与最多时间，与表格中实际到达时间相加得到。

在实际应用时，可结合航空公司自身情况代入实际数据进行求解，这样可得到更符合真实情况的结果。

5.3.3 航班排列优化模型的建立

5.3.3.1 模型一：不考虑延误状态的情形

(1) 目标函数

本题将以飞机场某一单跑道进行讨论。假设有 n 架飞机同时到达机场终端区，若想以最大限度避免飞机晚点情况的发生，必须使所有飞机着陆时间总和最小。因此，我们可写出如下表达式作为我们的目标函数：

$$\min T = \sum_{i=1}^n \sum_{j=1}^n T_{ij} x_{ij} \quad (13)$$

(2) 约束条件

由于飞机要尽可能的一架接连一架着陆才可能使所有飞机总着陆时间最小，固有如下公式的成立作为目标函数（式 13）的约束条件：

$$s.t \left\{ \begin{array}{l} \sum_{j=1}^n x_{ij} = 1, i=1,2,\dots,n \\ \sum_{i=1}^n x_{ij} = 1, j=1,2,\dots,n \\ x_{ij} \in \{0,1\}, i \neq j=1,2,\dots,n \end{array} \right. \quad (14)$$

其中,

T_{ij} : 飞机 i, j 之间必须保持最小时间间隔, 且 $T_{ij} \neq T_{ji}, i \neq j=1,2,\dots,n$

x_{ij} : 决策变量, $x_{ij} \in \{0,1\}$, 当 $x_{ij}=1$ 时, 表示飞机 j 紧跟飞机 i , 当 $x_{ij}=0$ 时表示飞机 i 紧跟飞机 j 。

此外, 还得保证飞机之间有足够的的时间间隔保证其安全。ICAO 规定了在无风条件下不同机型 (H: 重型, L: 大型, S: 小型) 之间尾流间隔的最小距离标准得到的最小安全时间间隔标准。

表 13 无风条件下最小安全时间间隔

	重	大	小
重	94	114	167
大	74	74	138
小	74	74	98

例: 第一行第二列的 114 表示当大型飞机 (L) 紧跟重型飞机着陆时, 之间的时间间隔最小为 114 分钟。

5.3.3.2 模型二: 考虑延误状态的情形

在不考虑延误状态下的模型中, 目标函数 (13) 仅单纯的考虑了飞机着陆完成时间最少这一因素。但在实际降落时, 飞机有可能出现延误, 并且着陆时管理员考虑的是综合效益, 所以更加符合实际情况的应是以总耗费最小体现在飞机着陆总时间上的变化为目标函数。因此, 我们可假设飞机拥有一个理想调度到达时间 t_j 满足如下关系:

(1) 若 $E_j \geq t_j$, 则飞机 j 的理想调度时间为 $STA_j = STA_i + T_{ij} + (E_j - t_j) = E_j$;

(2) 若 $E_j \leq t_j$, 则飞机 j 的理想调度时间为 $STA_j = STA_i + T_{ij} = t_j$ 。

其中,

t_j : 飞机 j 的理想调度时间;

STA_i : 飞机 i 调度到达时间;

STA_j : 飞机 j 调度到达时间;

T_{ij} : 飞机 i, j 之间必须保持最小时间间隔, 且 $T_{ij} \neq T_{ji}, i \neq j = 1, 2, \dots, n$;

E_j : 飞机 j 的最早到达时间。

结合飞机在空中延误要花费额外的延误时间的情况, 故改进后的模型应是:

$$\begin{aligned} \min \quad & T = \sum_{i=1}^n \sum_{j=1}^n T_{ij} \cdot x_{ij} + \sum_{j=2}^n \max\{E_j - t_j, 0\} + \alpha \sum_{j=2}^n \max\{t_j - L_j, 0\} \\ \text{s.t.} \quad & \begin{cases} \sum_{j=1}^n x_{ij} = 1, & i = 1, 2, \dots, n \\ \sum_{i=1}^n x_{ij} = 1, & j = 1, 2, \dots, n \\ x_{ij} \in \{0, 1\}, & i \neq j = 1, 2, \dots, n \end{cases} \end{aligned} \quad (15)$$

同时, 还应满足飞机间的安全间隔时间 (如表 13 所示)。

其中,

T_{ij} : 飞机 i, j 之间必须保持最小时间间隔, 且 $T_{ij} \neq T_{ji}, i \neq j = 1, 2, \dots, n$;

x_{ij} : 决策变量, $x_{ij} \in \{0, 1\}$, 当 $x_{ij} = 1$ 时, 表示飞机 j 紧跟飞机 i , 当 $x_{ij} = 0$ 时, 表示飞机 i 紧跟飞机 j ;

E_j : 飞机 j 最早到达时间;

L_j : 飞机 j 最晚到达时间;

t_j : 飞机 j 的理想调度时间。

我们将考虑延误状态的航班排列模型 (15) 作为我们的最终模型。

5.3.4 航班排列模型的求解

5.3.4.1 算法原理

由于该模型按传统方法求解较为困难, 而现有的一些求解方法有不少是基于传统的遗传算法进行求解。但是该方法局部搜索能力不强, 导致求出的解质量并非很高。所以, 我们考虑使用局部搜索能力强的模拟退火算法与其结合, 利用该混合算法来求解飞机排序的优化问题。这样既可以丰富求解的搜索范围, 也能保

证算法收敛。该算法步骤[7-9]如下，流程图如图 8 所示：

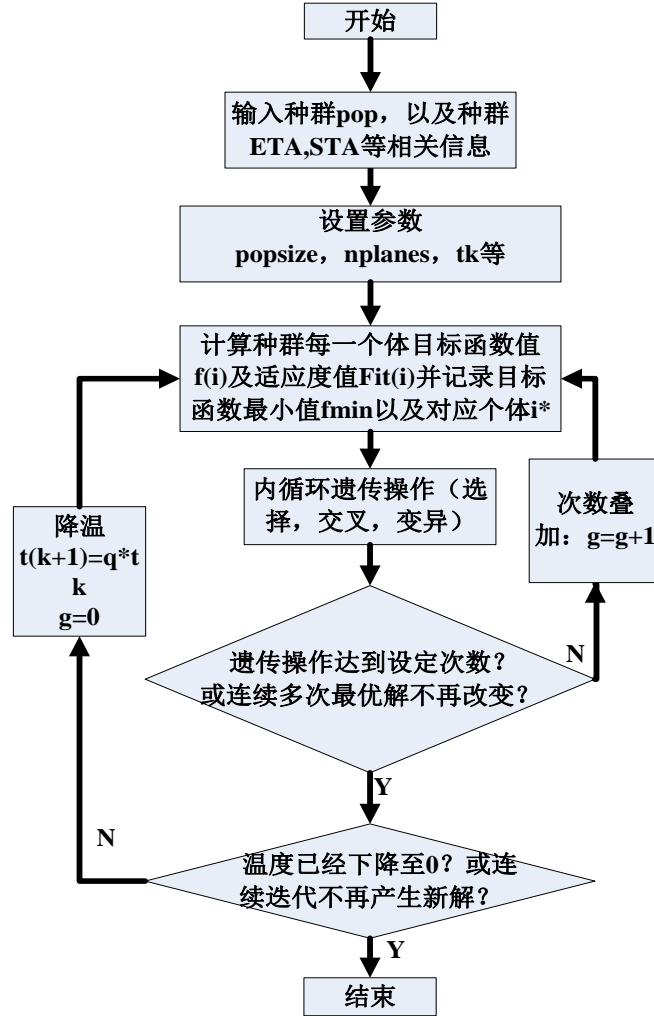


图 8 模拟退火遗传算法思路图

Step1: 设置参数 $popsiz$ 、 $nplanes$ 、 tk 、 pc 、 pm 、 $\max Gen$ 最大遗传代数)。初始化 $t = t_0, g = 0$, 随机生成初始种群 pop 。

Step2: 计算种群大每一个体 i 的目标函数值 $f_{(i)}$ ，及其适应度值 $Fit_{(i)}$ 。记录当代种群的目标函数最小值 $f_{\min} = f_{(i)}$ ，并把对应个体记为 i^* 。

Step3: 进行具有最优保存策略的选择操作，将父代种群大的最优个体直接复制到子代，子代大其余个体通过父代适应度，按选择概率进行选择复制，得倒新种群 $pops$ 。

Step4: 对选择后的新种群 $pops$ 进行 $Grefenstette$ 编码，并按一定的交叉概率 pc 进行交叉操作，完成交叉操作后，再对个体进行编码，计算新个体的目标

函数值，同时利用模拟退火算法的 Metropolis 准则判别新个体是否被接受，最终生成交叉之后的新种群 $popx$ 。

Step5: 对交叉操作后的新种群 $popx$ 按一定的变异概率 pm 进行变异操作，对变异后的新个体计算目标函数值，也要利用模拟退火算法的 Metropolis 准则判别新个体是否被接受，最终生成变异之后的新种群 $popm$ ，即形成新一代种群 $newpop$ 。

Step6: 降低温度，回到 Step2，继续。

5.3.4.1 算法特殊设计

本文所使用的模拟退火遗传算法，在遗传算法部分对一些内容做了适当改进以适应该模型求解。

传统遗传算法的适应度函数一般沿用目标函数进行运算。但在本模型中，目标函数求得的值是所有飞机着陆所花总时间最小值，而适应度函数值要体现的是种群中优良个体的较高适应度值。故本模型进行求解时，不能使用原目标函数来作为适应度函数。经查阅相关文献[8]结合传统遗传算法中适应度函数的特点，我们设计如下适应度函数：

$$Fitness(f(x)) = \exp\left(-\frac{f(x) - \min f(x)}{t_k}\right) \quad (16)$$

其中，

$f(x)$: 染色体对应目标函数值

$\min f(x)$: 当代进化群中最小的目标函数值

t_k : 温度参数

在该适应度函数中，当温度 t_k 较高时，目标函数值与种群中最小目标函数值之差较小，故在高温状态目标函数值大的个体适应度与目标函数值小的个体适应度相差不大。而当温度 t_k 较低时，目标函数值与种群中最小目标函数值之差变大，使得个体适应度差距得到充分体现。因此，该适应度函数可以提高算法收敛的速度，还可以体现高温粗搜索，低温细搜索的特点，利于寻找最优解。

5.3.4.3 求解结果

根据文献[9]，设置基本参数：交叉概率 $pc=0.7$ ；变异概率 $pm=0.1$ ；终止代数为 200 代； $popsize$ 的初始值为 30。

然后进行计算，运用 MATLAB 2009a（程序见附录 3）求得如下结果：

表 14 深圳宝安机场 2015 年 5 月 12 日部分航班最优调度时间

排列	计算结果	着陆时间	着陆总耗时	模拟退火后最终温度
9	1448	14:48:00		
1	1449.3	14:49:18		
8	1450.9	14:50:54		
2	1452	14:52:00		
7	1453.4	14:53:36	11.50	2.4179
10	1454.9	14:54:54	分钟	
4	1456.1	14:56:06		
3	1457.2	14:57:12		
5	1458.4	14:58:24		
6	1459.5	14:59:30		

从表 14 的结果可以看出，模拟退火后的最终温度为 2.4179，能够达到比较好的优化效果。

5.3.4.2 结果分析

(1) 由表 14 可知，经程序运行后我们可得到这 10 架飞机的最佳排列顺序，但该排列顺序是由随机生成的 10 架飞机任意排列后的种群中产生，且各飞机存在延误的可能性，所以仅代表该种群下的最优解。该排列顺序耗时 **11.50** 分钟。此外，我们列出了该排列顺序下每架飞机的最优调度时间，有利于机场工作人员的协调调度。

(2) 离港与着陆情形相同，仅需代入离港相关数据即可得出相应的飞机排列顺序及理想调度时间，在此不再赘述。

(3) 该算法优点为计算速度快，结果可靠，局部搜索能力强。但由于本文所采用的方法为随机生成 30 个 1 至 10 排列的数组的方法求满足目标函数的最优排列顺序，并且在计算过程中，所利用的遗传算法部分中有交叉变异算子，使得过程具有较大随机性。此外，该算法中每次种群变化都会按照 Metropolis 准则判断新种群是否被接受，因此随机性更强。

综上所述，虽然程序每次运行的结果都不同，但最大程度保证了种群的丰富性，使得各种情况的出现都可以得到充分考虑。若扩大初始种群数量并且进行多次运算后，可减小各过程随机性进而得到更符合实际情况的最优解。

6 模型的评价与推广

6.1 模型的优点

(1) 有效简便。采用 Matlab2009a 专业软件对模型进行求解，使运算更为简便快捷，效率更高；

(2) 结果可靠。本文同时使用熵权法与主成分分析法，能有效避免主观赋权造成的影响，使结果更加合理可靠。

(4) 算法适用范围广。本文所使用的模拟退火遗传算法局部搜索能力强，且适用于各类复杂优化问题的求解。

(3) 模型简单易懂，方法灵活，具有较强的推广性。

6.2 模型的缺点

由于本题所搜集的某些数据并非很充足，求解后得到的一些结论仍有待大数据量检验。若能提供更多详实数据，效果将更佳。

6.3 模型的推广

本文在建立航班排列模型大，能巧妙运用多种方法进行指标筛选，并创新的使用了模拟退火算法和遗传算法相结合的启发式搜索算法求解模型，可广泛用于路径规划，系统设计等实际问题大，有较强的推广性。

参考文献

- [1]FlightStats 官方网站,Global Cancellation and Delays<http://www.flightstats.com/go/Home/home.do>,2015/05/13
- [2]美国运输部, Airline On-Time Statistics<http://www.dot.gov>, 2015/05/13
- [3]中国民用航空局官网, 民航行业发展统计公报, <http://www.caac.gov.cn> ,2015/05/13
- [4]深圳宝安国际机场官网, 离港航班延误查询, <http://www.szairport.com>
- [5]章穗, 张梅, 迟国泰. 基于熵权法的科学技术评价模型及其实证研究[J]. 管理学报, 2010, 7:34-42.
- [6]李艳双, 曾珍香, 张闽,等. 主成分分析法在多指标综合评价方法中的应用[J]. 河北工业大学学报, 1999, 28(1):94-97.
- [7]王雪梅, 王义和. 模拟退火算法与遗传算法的结合[J]. 计算机学报, 1997, (4):381-384.
- [8]金芬, 孙春华, 钟鸣. 遗传算法中适应度函数的改进[J]. 机械设计与制造, 2010, (3):218-219.
- [9]李秋颖. 基于模拟退火遗传算法的机场终端区飞机排序问题的研究[D]. 河北工业大学, 2013.

附录（所有程序软件均为 Matlab 2009a）

1. 题目所给的判断是否正确

1.1 15min 标准下的航班延误率排序

```
clear all
clc
NUM=xlsread('第一问终极版.xls','C4:E18');
num=xlsread('第一问终极版.xls','B4:F18');
for n =1:15
r(n)=sum(num(n,:));
end
b=r
for j =1:15
t(j)=sum(NUM(j,:));
end
c=t
q=c./b;
digits(4)
ww=vpa(q)
sort(ww)
[c,ord]=sort(ww,2)
```

1.2 30min 标准下航班延误率排序

```
clear all
format short

syms njkrbcbq
[NUM]=xlsread('第一问终极版','D4:E18');
[num]=xlsread('第一问终极版','B4:F18');
for n =1:15
r(n)=sum(num(n,:));
end
b=r
for j =1:15
t(j)=sum(NUM(j,:));
end
c=t
q=c./b;
digits(4)
ww=vpa(q)
sort(ww)
[c,ord]=sort(ww,2)
```


1. 3 45min 标准下航班延误排序

```
clear all
clc
NUM=xlsread('第一问终极版','E4:E18');
num=xlsread('第一问终极版','B4:F18');
for n =1:15
r(n)=sum(num(n,:));
end
b=r
for j =1:15
t(j)=sum(NUM(j,:));
end
c=t
q=c./b;
digits(4)
ww=vpa(q)
sort(ww)
[c,ord]=sort(ww,2)
```

1.4 中美两国航班延误初步分析

```
clear all
clc

%%2011年%%
%%中国
a=xlsread('中国航班延误原因比例 (06-13)','B7:E7');
subplot(3,2,1)
pie(a)
title('2011年中国','fontsize',18)
legend('航空公司自身原因','流量控制','天气原因','其他')
%%美国
b=xlsread('美国航班延误原因比例','J5:J9');
subplot(3,2,2)
pie(b)
title('2011年美国','fontsize',18)
legend('Air Carrier Delay','Air Arriving Late','Security Delay','National Aviation System Delay','Extreme Weather')
%%2012年%%
%%中国
c=xlsread('中国航班延误原因比例 (06-13)','B8:E8');
subplot(3,2,3)
pie(c)
title('2012年中国','fontsize',18)
```

```

legend('航空公司自身原因','流量控制','天气原因','其他')
%%美国
d=xlsread('美国航班延误原因比例','K5:K9');
subplot(3,2,4)
pie(d)
title('2012年美国','fontsize',18)
legend('Air Carrier Delay','Air Arriving Late','Security Delay','National Aviation System Delay','Extreme Weather')
%%2013年
%%中国
e=xlsread('中国航班延误原因比例 (06-13)','B9:E9');
subplot(3,2,5)
A=pie(e)
title('2013年中国','fontsize',18)
legend('航空公司自身原因','流量控制','天气原因','其他')
%%美国
f=xlsread('美国航班延误原因比例','L5:L9');
subplot(3,2,6)
pie(f)
title('2013年美国','fontsize',18)
legend('Air Carrier Delay','Air Arriving Late','Security Delay','National Aviation System Delay','Extreme Weather')

```

1. 5 造成中国航班延误的因素分析

```

clear all
NUM=xlsread('中国航班延误原因比例 (06-13)','B2:E9')
A=NUM'
for i=1:4
a=A(i,:)
X=[ 2006 2007 2008 2009 2010 2011 2012 2013 ]    %%折线图中横向数据
Y=a
plot(X,Y,'k-o')    %%plot作图，“k-o”表示实线
ylabel('耽搁航班的比例','fontname','标楷体')
xlabel('年份','fontname','标楷体')
%%ylabel和xlabel是设置X和Y轴的名称、单位、字体

set(gca,'xtick',[ 2006 2007 2008 2009 2010 2011 2012 2013 ])
%%gca表示当前对象句柄，set命令对当前对象设置坐标轴xtick属性，
%%此属性表示了横坐标实际绘制的刻度范围，也可使用为
%%set(gca,'ytick',Y)
TITLE('2006年至2013年机场不同原因造成航班延迟的情况')
hold on%%第一条线完成

```

end

1. 6 造成美国航班延误的不同因素分析

```
clear all
[NUM]=xlsread('C:\MATLAB701\work\美国','B5:M9')
A=[NUM]
for i=1:5
a=A(i,:)
X=[2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014]    %%折线图中横向数据
Y=a
plot(X,Y,'k-o')    %%plot作图，“k-o”表示实线
ylabel('耽搁航班的比例','fontname','标楷体')
xlabel('年份','fontname','标楷体')
%%ylabel和xlabel是设置X和Y轴的名称、单位、字体

set(gca,'xtick',[2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014])
%%gca表示当前对象句柄，set命令对当前对象设置坐标轴xtick属性，
%%此属性表示了横坐标实际绘制的刻度范围，也可使用为
%%set(gca,'ytick',Y)
TITLE('2003年至2014年美国机场不同原因造成航班延迟的情况')
hold on%%第一条线完成

end
```

2. 我国航班延误主要原因

2.1 熵权分析法求解我国机场延误主要因素

```
clear
clc
x=xlsread('中国航班延误原因比例(06-13)','B2:E9');
a=min(x);
b=max(x);
[n,m]=size(x);
k=1/log(n);
for i=1:n
for j=1:m
x(i,j)=(x(i,j)-a(j))/(b(j)-a(j))*100;    %高优指标处理
end
end
for i=1:n
for j=1:m
x(i,j)=(b(j)-x(i,j))/(b(j)-a(j))*100;    %低优指标处理
end
end
```

```

end
he=sum(x);
for i=1:n
for j=1:m
    p(i,j)=x(i,j)/he(j);
end
end%指标归一化
for i=1:n
for j=1:m
if p(i,j)==0
    z(i,j)=0;
else
    z(i,j)=log(p(i,j))
end
end
end
e=zeros(1,m);
for i=1:ln
for j=1:m
    e(j)=e(j)+p(i,j)*z(i,j)*(-k);
end
end
he=sum(e);
for i=1:m
    g(i)=1-e(i)/(m-he) ;
end
for i=1:m
    w(i)=g(i)/sum(g); %计算权重
end
s=zeros(1,n);
for i=1:n %计算综合得分
for j=1:m
    s(i)=s(i)+w(j)*p(i,j);
end
end
s

```

2.2 主成分分析下中国航班延误主要贡献率

```

clear all
dataset=xlsread('航班延误表格(06-13)', 'B2:E9')
stdr=std(dataset); %求各变量的标准差
[n,m]=size(dataset); %定义矩阵行列数
sddata=dataset./stdr(ones(n,1),:); %将原始数据采集标准化
sddata %输出标准化数据

```

```

[p, princ, eigenvalue, t2]=princomp(sddata); %调用前三个主成分系数
p3=p(:, 1:3); %提取前三个主成分得分系数, 通过看行可以看
出对应的原始数据的列, 每个列在每个主成分的得分
p3 %输出前三个主成分得分系数
sc=princ(:, 1:3); %提取前三个主成分得分值
sc %输出前三个主成分得分值
e=eigenvalue(1:3)'; %提取前三个特征根并转置
M=e(ones(m, 1), :).^0.5; %输出前三个特征根并转置
compmat=p3.*M; %利用特征根构造变换矩阵
per=100*eigenvalue/sum(eigenvalue); %求出成分载荷矩阵的前三列
per %求出各主成分的贡献率
cumsum(per); %列出各主成分的累积贡献率

clf
A=[62.8192 29.9979 7.1825 0.0003]';
figure(1), bar([A, 0.5, 'stacked', 'linewidth', 3)
legend('A', 2)
set(gca, 'linewidth', 2.5, 'fontsize', 24, 'fontname', 'Arial')
set(gcf, 'color', 'w')
%title('bar/粗细0.5')

%title('bar/grouped')
%title('影响中国机场航班延误的各因素的贡献率')
figure(3), bar3([A, 'stacked'])
legend('A', 1)

set(gca, 'linewidth', 2.5, 'fontsize', 24, 'fontname', 'Arial')
set(gcf, 'color', 'w')

```

3. 基于模拟退火遗传方法的措施改善

```

%popsize 给定群体规模;
%pid 到场飞机固有编号;
%pxh 到场飞机固有型号;
%pjg 飞机尾流安区间隔矩阵;
%pop 初始种群;
%cpx 当代种群按目标函数值从小到大排序结果;
%cmin 当代目标函数最小值;
%Fit 个体适应度值;
%Fpx 适应度值按从大到小排序结果;
%best 最优排序队列编号;
%bestchromm 最优排序个体;
%pfrit 按适应度值排名后计算的个体轮盘赌选择概率;
%pops 选择操作后的新种群;

```

```

%grefpop 对选择操作后的种群进行Grefenstette编码;
%popx 在Grefenstette编码方式下进行交叉操作后的种群;
%popxjm 完成交叉操作后,进行Grefenstette解码;
%metpopx Metropolis准则判别后的交叉操作所得种群;
%popm 变异操作后的种群;
%newpop 变异操作后由Metropolis判别后的种群;
%pc 交叉概率;
%pm 变异概率;
%swappop 模拟退后扰动后种群;
%n 受扰动个体数量;
%sapop 由模拟退火降温后种群;
%initial_temperature 初始温度;
%tk 温度参数;
%iterations 遗传操作迭代次数;
%threshold 迭代次数控制参数;
% cooling_rate降温比率;
%complete_temperature_iterations 恒温下迭代次数;
%1、参数初始化、产生初始种群
global iterations;initial_temperature=1000; tk=initial_temperature;
nplanes=10;n=nplanes/2;iterations=1;complete_temperature_iterations=0;
threshold=1000;a=1.1;cooling_rate=0.8;pc=0.7;pm=0.1;
pid=1:nplanes;%飞机编号
pxh=xlsread('深圳宝安国际机场','B2:B11');%飞机型号
pjpg=[1.34,1.56,2.47;1.14,1.14,2.18;1.14,1.14,1.38]; %已将时间间隔矩阵换算为以分钟
    钟为单位的数字
for i=1:30
    pop(i,:)= randperm(10);
    i=i+1;
end
pop
popsize=30;
C2=zeros(30,1);
ETA=xlsread('深圳宝安国际机场','E2:E11');
STA=xlsread('深圳宝安国际机场','F2:F11');
E=xlsread('深圳宝安国际机场','G2:G11');
L=xlsread('深圳宝安国际机场','H2:H11');
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%主程序%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
for i=1:threshold
    u=size(pop,1);
    popsize=u;
    C=Objective2(popsize,nplanes,pxh,pjpg,pop,STA,E,L);
    C1=C2;
    C2=C;
    [fpx,Fit,Fpx,best,bestpop]=Fitness(popsize,pop,C,tk);

```

```

pops=select(popsiz, nplanes, pop, Fit, best);
grefpop=Grefcode(pops);
if popsiz>1
popx=crossover(popsiz, nplanes, pc, grefpop);
v=size(popx, 1);
popsiz=v;
else
break
end
popxjm=Grefdecode(popsiz, nplanes, popx, pid);
metpopx=crossMetropolis(popsiz, nplanes, popxjm, pops, pxh, pjg, tk);
[newpop, popm]=mutation(popsiz, nplanes, pm, metpopx, pxh, pjg, tk, popxjm);
swappop=swap(popsiz, nplanes, newpop, n);
iterations=iterations+1
if iterations<threshold %内循环终止条件
sapop=simulatedannealing(popsiz, nplanes, iterations, complete_temperature_iteration
s, cooling_rate, tk, threshold, swappop, C1, C2);
tk=cooling_rate*tk
else
break
end
if tk<=0 %外循环终止条件
break
end
sapop(all(sapop==0, 2), :)=[];
pop=sapop
end

```

3.2 目标函数值计算

3.2.1 不考虑时间窗及延迟的着陆总时间

```

function c=Objective1(popsiz, nplanes, pxh, pjg, pop)
for i=1:popsiz
s=0;
for j=2:nplanes;
s=s+pjg(pxh(pop(i, j-1)), pxh(pop(i, j))); %计算每一队列的降落总时间, 目标函数
end
c(i)=s;
end

```

3.2.2 考虑时间窗及延迟的着陆总时间

```

function C=Objective2(popsiz, nplanes, pxh, pjg, pop, STA, E, L)
for i=1:popsiz
t(i, 1)=STA(pop(i, 1));
for j=2:nplanes
t(i, j)=t(i, j-1)+pjg(pxh(pop(i, j-1)), pxh(pop(i, j)));
end

```

```

end
for i=1:popsize
for j=1:nplanes
if t(i, j)<=E(pop(i, j))
s(i, j)=E(pop(i, j)); %s(i, j)表示飞机最终调度后到达时间
else
s(i, j)=t(i, j);
end
end
cost(i, 1)=0;
for j=2:nplanes
if t(i, j)<=L(pop(i, j))
cost(i, j)=s(i, j)-s(i, j-1);%cost代表所花费总时间
else
cost(i, j)=s(i, j)-s(i, j-1)+t(i, j)-L(pop(i, j));
end
end
end
C=sum(cost, 2);%cost行求和
s
t

```

3.3 适应度函数

```

function[fpx, Fit, Fpx, best, bestpop]=Fitness(popsize, pop, C, tk)
fpx=sort(C); %将当代种群中的目标函数从小到大排列
fmin=fpx(1); %记录当代种群中的目标函数最小值
best=find(C==fmin); %记录当代种群中的目标函数最小值的最优个体
bestpop=pop(best, :); %记录当代种群中最优个体对应具体排列方案
for i=1:popsize
Fit(i)=exp(-(C(i)-fmin)/tk); %利用模拟退火机制计算适应度
end

```

```

Fpx=sort(Fit, 'descend'); %为后面的选择过程做准备

```

3.4 用模拟退火遗传算法产生新种群

3.4.1 选择操作，轮盘赌选择算子

```

function pops=select(popsize, nplanes, pop, Fit, best)
pfit=Fit./sum(Fit);%pfit 按适应度值排名后计算的个体轮盘赌选择概率;
psfit=cumsum(pfit); % cumsum函数通常用于计算一个数组各行的累加值
if psfit(1, popsize)<1
psfit(1, popsize)=1;
end
rs=rand(1, popsize); %生成一个1*popsize的随机数组
l=length(best);
for i=1:l
pops(i, :)=pop(best(i), :);
%实行最优保存策略, 将上一代适应度值最高的个体直接复制到下一代, 同时考虑适应度值相等

```


的飞机排列方式可能会有多个, 将多种方案全部保留

```
end
for i=1+1:popsize
    ss=0;
    for j=1:popsize
        if rs(i)<=psfit(j)
            pops(i,:)=pop(j,:);
            ss=1;
        end
    end
    if ss==1
        break;
    end
end
end
```

3. 4. 2 交叉操作

%%由于飞机队列中的飞机不能出现重复, 故对进行了选择操作的新群体首先使用格里芬斯塔特编码方案进行编码, 然后进行单点交叉, 并在生成新种群后, 在对新种群进行解码, 并利用 Metropolis 准则判别新个体是否被接受.

3. 4. 2. 1 编码

```
function grefpop=Grefcode(pops)
[popsize,nplanes]=size(pops);
for i=1:popsize
    a=1:nplanes;
    for j=1:nplanes
        x=pops(i,j);
        grefpop(i,j)=find(a==x);
    end
    y=grefpop(i,j);
    a(:,y)=[];
end
end
```

3. 4. 2. 2 交叉操作

```
function popx=crossover(popsize,nplanes,pc,grefpop)
for i=1:popsize/2
    if rand<pc
        crossw=fix(rand*(nplanes-1))+1;%crossw为交叉点, fix截尾取整函数
        p1=grefpop((2*i-1),:);
        p2=grefpop(2*i,:);
        popx((2*i-1),:)=p1(1:crossw),p2((crossw+1):nplanes)];
        popx(2*i,:)=p2(1:crossw),p1((crossw+1):nplanes)];
    else
        popx((2*i-1),:)=grefpop((2*i-1),:);
        popx(2*i,:)=grefpop(2*i,:);
    end
end
end
```

3.4.2.3 解码在进行过交叉操作后对所得新种群进行解码，回应到飞机编号

```
function popxjm=Grefdecode(popsiz, nplanes, popx, pid)
for i=1:popsiz
    tempw=pid;
for j=1:nplanes
    popxjm(i, j)=tempw(popx(i, j));
for k=(popx(i, j):(nplanes-j)
    tempw(k)=tempw(k+1);
end
end
end
```

3.4.2.4 Metropolis 准则判别

```
function metpopx=crossMetropolis(popsiz, nplanes, popxjm, pops, pxh, pjg, tk)
for i=1:popsiz
    e1=0; e2=0;
for j=2:nplanes
    e1=e1+pjg(pxh(popxjm(i, j-1)), pxh(popxjm(i, j)));
%计算交叉后新队列的降落总时间——目标函数（即是能量函数）
    e2=e2+pjg(pxh(pops(i, j-1)), pxh(pops(i, j)));
%计算未交叉的选择操作后的队列的降落总时间——目标函数（即是能量函数）
    if e1<=e2
        metpopx(i, :)=popxjm(i, :);
%在交叉后队列能量低于原队列时，接受交叉后新队列
    else
        if rand<exp(-(e2-e1)/tk) %在交叉后队列能量高于原队列时，按Metropolis准则接受交叉后
            新队列
            metpopx(i, :)=popxjm(i, :);
        else
            metpopx(i, :)=pops(i, :);
        end
    end
end
end
```

3.5 变异

```
function [newpop, popm]=mutation(popsiz, nplanes, pm, metpopx, pxh, pjg, tk, popxjm)
popm=metpopx;
for i=1:popsiz
    x=fix(rand*nplanes)+1;
    y=fix(rand*nplanes)+1;
    if rand<pm
        popm(i, x)=popxjm(i, y);
        popm(i, y)=popxjm(i, x);
    end
end
e1=0; e2=0;
```

```

for j=2:nplanes
e1=e1+pjg(pxh(popm(i, j-1)), pxh(popm(i, j)));%计算变异后新队列的降落总时间——能量函数
e2=e2+pjg(pxh(metpopx(i, j-1)), pxh(metpopx(i, j)));%计算未变异交叉操作后的队列的降落总时间——能量函数
if e1<=e2
newpop(i,:)=popm(i,:);%在变异后队列能量低于原队列时，接受变异后新队列
else
if rand<exp(-(e2-e1)/tk)%在变异后队列能量高于原队列时，按Metropolis准则接受变异后新队列
newpop(i,:)=popm(i,:);
else
newpop(i,:)=metpopx(i,:);
end
end
end
end

```