

航班恢复问题

1. 背景

随着经济的发展，航空出行已成为越来越多旅客的选择。但众所周知，飞机航班如果不能按原计划执行，不仅会给航空公司造成巨大的经济损失，同时还会给旅客出行带来极大的不便。在造成航班不正常的种种因素中，有些是不可抗拒的自然因素，如暴风雪、飓风等，有些是不可预测的突发事件，如突发恐怖袭击、飞机机械故障等等，还有些是因为管理手段的落后，比如飞行员缺位、空中管制，等等。下表是 FlightStats 网站公布的今年二月份世界主要航空公司和部分中国航空公司航班准点率的比较。可以看出，虽然中国的航班准点率很低，但其他国家和地区也不乐观，比如美国本土的平均航班准点率也只有 77%。

航空公司	名次	准点率%	航空公司	名次	准点率%
Iberia	1	92.45	United (美联航)	19	81.99
Singapore (新航)	2	88.14%	Cathy Pacific (国泰)	30	75.03
Delta (美三角)	3	87.54	Air China (国航)	38	66.55
American (美航)	6	86.2	China Eastern (东航)	39	61.74

需要指出的是，由于目前中国航空公司在国内主要航线上航班安排已经比较稠密，一旦某个航班出现故障，就有可能造成一系列的连锁反应，影响成千上万旅客的出行。一些航空公司没有把航班延误作为要事来抓，缺乏有效应对手段。如果抱着“等着瞧”的消极态度，不仅可能造成更多的没有必要的延误，而且还会导致最终产生一个失败的决策。例如航空公司在等待 3 个小时后，最终决定取消该航班，部分旅客被安置到此后 2 小时以后的某航班上。这样的结局显然不如一开始就宣布取消该航班，把旅客延迟到某航班上。

世界范围内，近年来快速增长的航空旅客数量已超过了很多人主要机场的容量，加上近年气候的反常变化和安全突发事件的增多，航班恢复问题越来越受到各国民航管理机构和各大航空公司的重视，中国主要航空公司也已经把航班恢复的自动化提到了议事日程上了。

最近发生的美国联航乘客被打事件，表面上是一个旅客服务管理问题，但本质上是航班恢复管理不慎造成的结果。联航为了避免外地航班机组人员缺位，紧急从芝加

哥基地调遣机组前往。由于机组缺位造成的航班中断有扩散到整个网络的可能，联航赋予了他们很高的登机优先级。这些都是正确的决策并且被正确地执行了，但在最后环节，联航工作人员没有能把座位“拍卖”坚持到最后时刻，从而导致了世界民航史上的这一重大事件的发生，给联航造成了不可挽回的重大损失。

其实，航班恢复问题的“难”除了相关因素的复杂，更主要的原因在于恢复方案的即时性。航班紊乱发生后，恢复方案的决定和实施是越早越好。在手工调整的情况下，调度员只能考虑到影响飞行安全的一些基本因素，很难考虑到全局网络的优化，更别说旅客的行程规划或根据旅客价值信息确定航班恢复的优先级了。举个最简单的例子，假如飞行网络中有一架飞机出现故障需要检修，受影响的航班可能不超过 10 个，具有多年调度经验的调度员大概需要几十分钟甚至 1~2 小时进行航班手工调整。可以想象，如果飞行网络出现大面积紊乱，受影响的航班可能有几十个甚至上百个，期望调度员手工在十几分钟甚至几分钟内完成整个网络的调整一定是异想天开，但借助于计算机求解数学优化模型却是可行的。

要最终可行，还有两个关键因素必须解决：1. 如何创建合适的数学模型；2. 如何用合适的算法快速求解这个数学模型。学术界研究航班恢复问题已经很久，取得了很好的进展，但业界至今还很少有实际的应用解决方案。因为理论研究一般都局限于有限的时间和空间，运行约束也仅仅是实际约束的部分子集，这样的方法很难被航空公司的运控部门采纳而直接用于生产实践。目前世界上提供解决航班恢复问题的产品寥寥无几，具有完整功能、满足各种实际需求的产品还只有 Sabre 一家。本赛题就是针对以上这两个问题而设计的。Sabre 公司通过在高校开展学术竞赛来提高学术界对不正常航班的恢复研究的关注度。更多相关的资料可以在以下网页浏览下载，

<http://page.sabreairlinesolutions.com/orcompetitions>。

总而言之，创建合适的数学模型和采用行之有效的算法求解是解决航班恢复问题的关键。目前，学术界一般采用 decomposition（例如 Bender's Decomposition 或者 Column Generation）的方法来求解这一类整数规划模型【1】【4】，更好的算法还有待于发现。

2. 问题介绍

航班恢复问题本质上是运营恢复问题的一部分。或者说，广义的航班恢复就是运营恢复，包括（狭义的）航班恢复（Flight Recovery）、机组恢复（Crew Recovery）和

旅客行程重新规划（Passenger Re-accommodation）三部分，它们相互约束，构成一个整体上超大规模的运筹优化问题。这个优化问题具有难以想象的复杂度，不是工业界目前已有计算机的计算能力所及。在实际运营过程中，航空公司是按流程次序先考虑航班恢复，然后在此基础上机组恢复，最后重新规划旅客的行程，把他们送往各自的目的地。相应地，采用运筹优化方法解决运营恢复问题也是按这三步把整个大问题按阶段次序分解成子问题【1】，即首先求解航班恢复问题，在此基础上求解机组恢复问题和旅客行程再规划问题。

需要指出的是，由于缺少信息交互，虽然每个子问题的求解可以达到局部最优，但整体最优却得不到保证，甚至有出现不可行解的可能。已经有学者证明，整合两个或者三个子问题成一个单一数学模型，可以得到更好质量的解【4】。所以本赛题作为航班恢复问题由四个子题目组成，从最基本的单一机型的航班恢复，多机型恢复，最后到考虑旅客行程重新规划的航班恢复。为了避免过于复杂化，本赛题不考虑机组人员的恢复，也不考虑旅客行程重新规划。

针对航班恢复问题，通常有三种航班调整方法：航班延误、飞机置换和航班取消。航班延误和飞机置换可以同时发生。

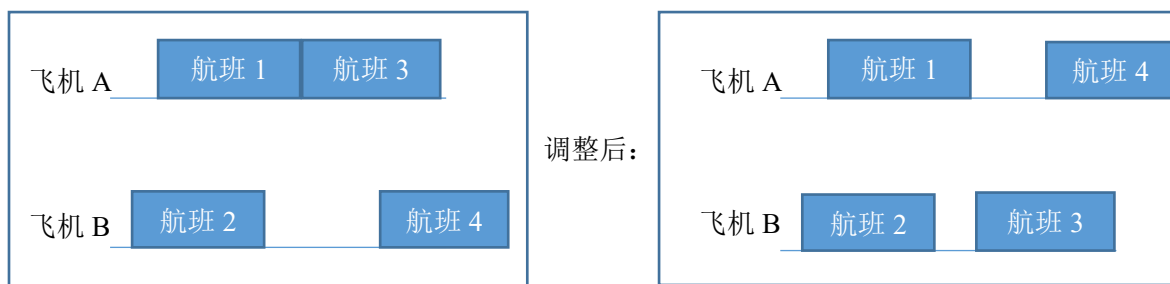
航班延误

如果选择航班延误，还需要给出具体的延误时间（以分钟为单位）。通常航空公司对延误都有最大延误时间约束，本赛题规定航班最大延误时间为 5 小时，即延误超过 5 小时时一定取消该航班。以间隔 10 分钟为一个决策单位，那么一个航班就有 30 个延误决策可选。航班延误的代价除了旅客满意度降低外，更重要的是联程旅客可能赶不上下趟航班。

本赛题规定，飞机的飞行时间不会因延误而受影响。

飞机置换

飞机置换就是将航班安排给不同于原计划执行飞机的其他飞机去执行。如下图所示，按计划，航班 1 连接航班 3 由飞机 A 执行，航班 2 连接航班 4 由飞机 B 执行。但由于延误，飞机 A 执行完航班 1 后没有足够时间间隔，无法及时执行航班 3。于是，调度员将航班 3 安排给飞机 B，航班 4 安排给飞机 A 去执行。



图一：飞机置换

飞机置换并不需要在完全相同的飞机之间进行，航空公司可以安排给满足约束条件的任何其它飞机。实际操作中，一般安排给同机型家族（比如 A320），或者同子机型（比如 A320-200）的任何一架飞机。飞机置换通常是最佳航班恢复方案，只要能满足最小飞机间隔时间就行。但这种机会不总是存在。

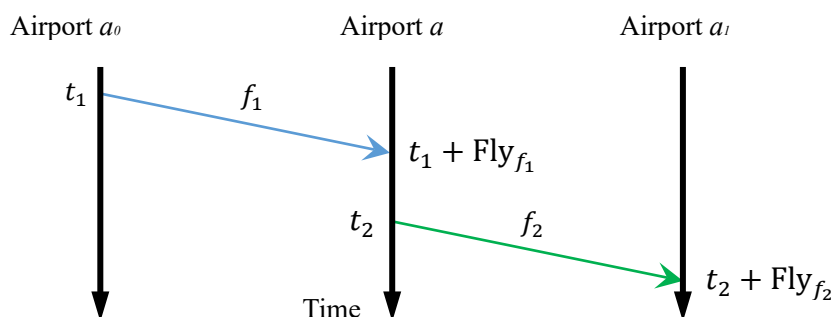
飞机间隔时间是指同一架飞机在执行完上一趟航班到执行下一趟航班前的地面停留时间。本题规定最小飞机间隔时间是 45 分钟。

航班取消

众所周知航班取消的含义，这里就省略解释。航班取消的代价显然是最严重的。

3. 数学模型示例

下面给出一个充分简化了的航班恢复问题的线性规划模型，供参赛者理解本赛题。参赛者可以在本模型的基础上完善并引入本赛题的具体目标和约束，比如最小飞机间隔时间，恢复期开始时航班的衔接，等等。但需要指出的是，建模方法很多（比



图二：时空网络（符号参照数学模型）

如参考文献里模型就不一样), 针对同一数学模型的算法也可以很多, 具体采用什么模型和算法, 以期在算法复杂度、解的质量和模型简洁方面达到平衡, 由参赛者自己决定。

最小化	$\sum_{f \in F} \text{CancelCost}_f \cdot (1 - x_f) + \sum_{f \in F} \text{DelayCost}_f \cdot \sum_{t \in T_f} (t - \text{Dpt}_f) \cdot x_{ft}$
约束条件	$x_f = \sum_{t \in T_f} x_{ft}, \forall f \in F$ $x_{ft} = \sum_{p \in P} x_{f tp}, \forall t \in T, \forall f \in F$ $y_{atp} = \sum_{f \in F_a^I} \sum_{\text{Dpt}_f \leq s < t - \text{Fly}_f} x_{f sp} - \sum_{f \in F_a^O} \sum_{\text{Dpt}_f \leq s < t} x_{f sp}, \forall a \in A, \forall t \in T, \forall p \in P, s \in T_f$
变量定义	x_f : 0-1 变量, 用来表示航班 f 是否运营. x_{ft} : 0-1 变量, 用来表示航班 f 是否在时间点 t 起飞. $x_{f tp}$: 0-1 变量, 用来表示航班 f 是否在时间点 t 由飞机 p 起飞. y_{atp} : 0-1 变量, 用来表示飞机 p 在时间点 t 是否停在机场 a .
参数符号	T : 时间轴上所考虑时间点的集合 F : 所有航班的集合 A : 所有机场的集合 P : 所有飞机的集合 $T_f \subseteq T$: 航班 f 的所有允许起飞时间点集合. Dpt_f : 航班 f 的最早允许起飞时间点, $\text{Dpt}_f = \min\{s: s \in T_f\}$. Fly_f : 航班 f 的飞行时间 $F_a^I \subseteq F$: 所有到达机场为 a 的航班集合. $F_a^O \subseteq F$: 所有起飞机场为 a 的航班集合. CancelCost_f : 取消航班 f 的成本. DelayCost_f : 航班 f 的每分钟的延误成本.

下面举例说明这个线性规划模型的复杂度。假设一个拥有 $|P| = 100$ 架飞机的子机型网络有 600 个航班需要执行。最大延误时间 5 小时, 每个航班有 $|T_f| = 30$ 个延误选择, 每个延误选择可以安排给 100 架飞机里的任意一架。这样大致估计共有 $|F| \cdot |T_f| \cdot |P| = 600 * 30 * 100 = 1.8 * 10^6$ 个选择变量 $x_{f tp}$ 。如果把这样一个有几百万个选择变量的 0-1 整数规划问题采用商用求解器 (如 CPLEX 或 GUROBI) 直接求解, 绝不可能保证在允许的时间限制内求得最优解。实际上, 这样的问题完全有可能运行几天甚或几年才能结束。如果我们的航班恢复问题考虑到多个子机型, 或者将 10 分钟的延

误决策间隔降低至 1 分钟（完全灵活地调整延误时间），这个问题将变得更加复杂。

除了上述时空网络的数学模型外，也可以将安排给某架飞机的所有航班考虑为一个有效路径，用列生成的方法去求解。这样可以减少因为延误时间而产生的大量决策变量【2】。

4. 赛题

本赛题有如下几个要求和规定：

1. 附件里时间的表达均为 Unix 格式，即从 1970 年 1 月 1 日 0 时 0 分 0 秒算起的秒数。从 Unix 时间格式到一般时间格式的转换可参考附件中 Excel 表格里的转换公式，比如 $((B2/60)/60)/24 + \text{DATE}(1970,1,1)$ 。另外，所有时间均假设为 UTC 国际标准时间。
2. 所有航班只能延误，不能提前，最早起飞时间不能早于原计划的起飞时间。
3. 各航班的飞行时间是常量，即航班数据中的到达时间减去起飞时间。
4. 保证每架飞机的连续航班能首尾相连，即前一航班的到达机场与后一航班的起飞机场必须相同，而且前一航班到达时间与后一航班起飞时间之间的最小间隔时间为 45 分钟。
5. 所有飞机的第一个航班要满足如下两个条件：（1）航班的起飞机场与飞机的起点机场一致，（2）航班的起飞时间不早于飞机的最早可用时间。
6. 所有飞机的最后一个航班的到达时间不能晚于飞机的最晚可用时间。
7. 航班延误的决策时间点间隔为 10 分钟（比如，安排给 X 飞机并按计划起飞，安排给 X 飞机并延误 10 分钟，安排给 X 飞机并延误 20 分钟，安排给 X 飞机并延误 30 分钟，以此类推）。如果参赛者有能力通过其他的数学模型找出更灵活的延误时间可以不考虑这一假设。
8. 不考虑机场可停留飞机的容量。理论上所有机场可以全天 24 小时工作。
9. 同行旅客是指一起订票并且行程完全一致的旅客，他们共享同一个旅客号，并作为一个整体考虑，即不能分乘不同的航班。
10. 所有航班，包括机场 OVS 关闭时段内的航班，它们的延误时间都需要被考虑到。为了最大可能保护航班，尽量不取消航班。
11. 数据结果提交的格式要求：参赛者除了在论文中必须叙述如何解读结果外，还

必须提交完整的新航班计划并作为附件与论文一起发送指定邮箱。提交的新航班计划有如下格式要求：

- (1) 虽然输入数据的时间都是 Unix 格式，所有输出数据的时间必须都是一般格式，比如，4/22/16 16:15 表示 2016 年 4 月 22 日 16 时 15 分。
- (2) 可以在原航班数据（见后）格式的基础上，添加“新机型号”、“新飞机尾号”、“新起飞时间”、“新到达时间”和“延误时间”。列的次序按下表示例。如有需要，参赛者还可在后面添加新列。新起飞时间和新到达时间与原时间如有不同，请用亮色标出。
- (3) 将问题一、问题二、问题三、问题四的新航班计划结果分别放在输出 Excel 总表格的 Sheet1、Sheet2、Sheet3、Sheet4 中。该 Excel 总表格的文件名为“C 参赛队号.xls”。表格各页的格式如下示例：

航班 唯一 编号	起飞 时间	新起 飞时 间	到达 时间	新到 达时 间	起飞 机场	到达 机场	飞机 型号	新飞 机型 号	飞机 尾号	新飞 机尾 号	延误 时间

- (4) 特别注意在该 Excel 总表格除外部文件名外，表格内容中不能有参赛队的任何信息，否则会当废卷处理。

另外需要指出，虽然好的算法设计应该考虑到输入数据的各种不同可能性，本赛题局限于求解附件里的具体案例。

由于受到暴风雪的影响，管理部门决定在 2016 年 4 月 22 日的 18:00 到 21:00 之间关闭机场 OVS。在该时间段内该机场不能起飞或降落任何航班，而该时间段之前的所有航班都处于正常状态，该时间段之后机场可立即恢复正常起降。因此，原定在该日 18:00 至 21:00 之间（不包括 18:00 和 21:00 这两个时刻）起降的所有航班都需要重新安排，而且它们的重新安排可能造成关闭后其它航班的重新安排。由于 OVS 机场的跑道限制，该机场每 5 分钟最多能起飞 5 架飞机，同时降落 5 架飞机。例如在 21:00 - 21:05 之间（不包括 21:05 时间点）最多有 5 个起飞航班和 5 个降落航班，其起飞和降落时间都可以按 21:00 计算；在 21:05 - 21:10 之间（不包括 21:10 时间点）最多有 5 个起飞航班和 5 个降落航班，其起飞和降落时间都可以按 21:05 计算；以此类推。其它机场不需要考虑跑道限制。

附件中给出了该航空公司在相关时间段内所有航班计划、飞机和旅客信息。摘录部分如下以便理解。

航班数据（Schedules）示例

航班唯一编号	起飞时间	到达时间	起飞机场	到达机场	飞机型号	飞机尾号
174774150	1461341700	1461348120	OVS	LEH	9	41098	
174774124	1461351000	1461356760	LEH	OVS	9	41098	
174777506	1461393900	1461399900	OVS	FUK	9	41098	
174777482	1461403200	1461409500	FUK	OVS	9	41098	
174777574	1461413400	1461422400	OVS	KMM	9	41098	
174773996	1461325560	1461335280	GAZ	OVS	9	51098	
...	

飞机数据（Aircrafts）示例

飞机尾号	飞机型号	最早可用时间	最晚可用时间	起点机场	座位数
41098	9	1461333600	1461426000	OVS	87	
51098	9	1461325560	1461426000	GAZ	87	
32098	9	1461330660	1461445800	VOR	87	
42098	9	1461302220	1461424200	OVS	87	
82098	9	1461326400	1461428400	OVS	87	
23098	9	1461330000	1461444600	QSM	87	
14098	9	1461326820	1461428700	OVS	87	
44098	9	1461325620	1461423900	OVS	87	
64098	9	1461323880	1461482400	OVS	87	
15098	9	1461326220	1461510000	OVS	87	
75098	9	1461323700	1461445800	OVS	87	
...	

旅客数据（Paxinfo）示例

旅客号	航班唯一编号	同行旅客数量
1	174777880	14
1	174781120	14
2	174773867	7
2	174778110	7
3	174778406	10
3	174777718	10
4	174777706	11

4	174777942	11
...

如果一个旅客号对应多个记录表示该旅客号搭乘多个联程航班，例如旅客号 1 搭乘了航班 174777880 和航班 174781120。如果一个旅客号只对应一个记录则表示该旅客号只搭乘一个直飞航班。该数据只用于求解第四问。

流控数据（Airportclosures）示例

机场	关闭开始时间	关闭结束时间
OVS	1461348000	1461358800

第一问：不考虑旅客信息，如何重新规划机型 9（不考虑其他机型）的航班计划，制定起飞时间表（给出延误分钟），使得所有原计划安排给机型 9 的航班尽可能不被取消，同时保证机型 9 的所有航班总体延误时间最短？此题可参考文献【3】。

第二问：不考虑旅客信息，假定同一机型的所有飞机的载客量相同，其间航班调整没有成本，但在不同机型间调整有成本。比方说飞机 DIBPV 属于 320 机型，飞机 COBPV 属于 321 机型，航班 174774110 原计划是安排给飞机 DIBPV 执行，如果将 174774110 分配给飞机 COBPV 执行则需要产生额外的成本。假设此额外成本等价于航班延误半小时（置换和延误有可能会同时发生，则成本叠加）。在这样的假设下如何重新规划飞机航班（包括所有机型的所有航班），制定起飞时间表（给出延误分钟）使原计划航班尽可能不被取消，同时保证所有航班总体延误时间最短？

第三问：进一步考虑飞机的载客量，假设在不同机型间调整航班的成本除了航班本身延误半小时外，还要加上不能登机旅客的成本（这里仍不考虑旅客的联程信息，即假定旅客行程都是直达的，并假设所有航班都是 100%的上座率）。比如飞机 DIBPV 的载客量是 140 人，COBPV 的载客量是 170 人。如果将飞机 COBPV 的航班分给 DIBPV 去执行，将会有 30 名旅客因没有座位而无法登机。但如果将 DIBPV 的原计划航班分配给 COBPV 去执行则没有这种情况。假设一名旅客无法登机与该旅客延误 2 小时的成本相当，该如何重新规划航班以保证旅客总体延误时间最短？

第四问：在第二题的基础上，假设在不同机型间调整航班不考虑成本。我们在旅客数据中提供了旅客的行程信息，包括旅客号，同行旅客数量，和相应的航班。每个旅客行程中的相连航班间最少需要 45 分钟间隔时间用于中转，如 23 日的航班 174778458（02:05 JOG—03:00 OVS）与 23 日的航班 174777524（05:50 OVS—08:10 XVS）的间隔时间为 2 小时 50 分钟。旅客的延误按照旅客计划到达最终目的地时间为基准计

算。例如在案例中旅客号为 6 的旅客计划到达 XVS 时间是 23 日 08:10，如果不晚于该时间到达则延误为 0，如果到达 XVS 时间是 23 日 08:40 则延误时间是 30 分钟，考虑旅客号为 6 的同行旅客数量为 8，则总体延误时间是 $8 \times 30 = 240$ 分钟。假定旅客号为 6 的旅客最终不能到达目的地相当于总体延误了 8×24 小时。该如何重新规划航班以保证旅客总体延误时间最短？如果某旅客号对应的航班号在航班表里找不到相应记录则不需要考虑该旅客。如果某航班没有对应的旅客信息，可认为该航班目前没有乘客，则延误该航班没有成本代价。

参考文献

- [1] Jon D. Petersen, Gustaf Sölveling, John-Paul Clarke, Ellis L. Johnson, and Sergey Shebalov (2012). An Optimization Approach to Airline Integrated Recovery. *Transportation Science*, 46(4), 482-500
- [2] J. M. Rosenberger, E. L. Johnson, G. L. Nemhauser(2003). Rerouting Aircraft for Airline Recovery. *Transportation Science*, 37(4), 408–421.
- [3] Ahmad I. Z. Jarrah, Gang Yu, Nirup Krishnamurthy, and Ananda Rakshit (1993). A Decision Support Framework for Airline Flight Cancellations and Delays. *Transportation Science*, 27(3), 266-280.
- [4] Stephen J. Maher (2015). Solving the Integrated Airline Recovery Problem Using Column-and-Row Generation. *Transportation Science*, 50(1), 216-237