

基于贝叶斯网络的航班延误波及研究

刘玉洁^{1,2},何丕廉¹,刘春波²,曹卫东²

LIU Yu-jie^{1,2},HE Pi-lian¹,LIU Chun-bo²,CAO Wei-dong²

1.天津大学 计算机科学与技术学院,天津 300072

2.中国民航大学 计算机学院,天津 300300

1.College of Computer Science and Technology,Tianjin University,Tianjin 300072,China

2.College of Computer Science and Technology,Civil Aviation University of China,Tianjin 300300,China

E-mail:tjliuyujie@msn.com

LIU Yu-jie,HE Pi-lian,LIU Chun-bo,et al.Flight delay propagation research based on Bayesian net.Computer Engineering and Applications,2008,44(17):242-245.

Abstract: In the industry of civil aviation,delay propagation of flight always is a main factor that impacts the flight's delay.This paper discusses the delay propagation problem in busy hub-airport.All the data used in the experiments is come from real records of one large domestic airlines.Through producing the model of delay propagation and modeling the Bayesian Net,the impact of arrival-delay on departure-delay between related flights is discussed.The outcomes of learned Bayesian net show that the existence of propagation phenomenon from arrival-delay impacts onto departure-delay;the degrees under varied circumstance are different;and the according coping methods are different.Cancellation of flights is a drastic remedy of those coping methods.It can weaken the delay propagation to some extent.The weakened degree is related to the number of flights cancelled.A preliminary reference can be made based on this research for airports when there is large scale delays of flights happened.

Key words: Bayesian net;delay propagation;arrival-delay;departure-delay;flight cancellation

摘要:在民航业内,航班延误波及问题一直是影响航班延误的一个主要因素。基于贝叶斯网络(BN),讨论了在繁忙的枢纽机场,其航班延误的波及问题。在实验中使用的数据,皆来自国内某大型航空公司的真实记录。通过建立延误波及模型和贝叶斯网络模型,探讨了相关航班中,进港延误和航班取消对离港延误的影响。学习的结果显示了进港延误(Arrival-delay)对离港延误(Departure-Delay)的波及现象的存在;以及波及现象在不同情况下的程度不同;相应的可采取的应对方式亦不相同。其中航班取消是一种釜底抽薪的应对方法,可以在一定程度上削弱上述条件下的延误波及,其削弱程度与航班取消的架次有关。基于该研究可以在机场发生大规模延误时,提供一个基本的参考。

关键词:贝叶斯网络;延误波及;进港延误;离港延误;航班取消

DOI:10.3778/j.issn.1002-8331.2008.17.072 **文章编号:**1002-8331(2008)17-0242-04 **文献标识码:**A **中图分类号:**TP301

1 引言

近年来随着经济的发展,我国的航空业也在迅速的腾飞,与此同时越来越显示出航班延误问题在制约着机场和航空公司的发展。对机场而言,航班延误严重影响机场的正常运作,对航空公司而言,频繁的航班延误不仅给航空公司带来了巨大的经济损失,更严重影响了航空公司的信誉。为此民航总局为提高民航的服务水平,近期也采取了多项措施,甚至强制取消了部分延误严重的航线。

由于航班延误上述的重要地位,它一直是国内、外航空业

一直致力研究的对象。国外对航班延误以及延误波及的研究起步较早,各类数据的记录也较国内详实得多,如 Ning Xu, George Donohue 等(Mason University)在其论文“Estimation of Delay Propagation in the National Aviation System Using Bayesian Networks^[1]”中,研究数据来自于联邦航空管理局(FAA)的航空系统性能度量 ASPM 数据库,通过建立贝叶斯网络模型,对美国三个繁忙的枢纽机场间的航班延误数据进行学习,分析了其中延误波及的影响。相对来讲,我国对航班延误的研究起步较晚,更多停留在管理学方面的损失评估和流量控制方面^[2]。

基金项目:国家高技术研究发展计划(863)(the National High-Tech Research and Development Plan of China under Grant No.2006AA12A106); 国家自然科学基金(the National Natural Science Foundation of China under Grant No.60572167)。

作者简介:刘玉洁(1973-),女,博士研究生,工程师,主要研究领域为神经网络、人工智能,民航信息化,嵌入式系统;何丕廉(1942-),男,教授,主要研究领域为人工智能与计算机辅助教育,自然语言处理,Web 检索与挖掘;刘春波(1976-),男,博士研究生,讲师,主要研究领域为虚拟现实,图像/视频编码;曹卫东(1964-),女,博士研究生,副教授,主要研究领域为数据库技术,数据挖掘。

收稿日期:2008-02-21 **修回日期:**2008-03-24

2 航班的延误波及模型及分析

2.1 航班的相关性

客观上来讲,航班延误是不可能彻底消除或避免的。航班的延误具备一个特殊性质——传播性。当一架航班由于各种原因产生延误时,它会引发连锁反应,将此次延误传播到与其相关的所有下游航班、甚至下游机场的其他航班,称这种现象为延误波及^[3]。根据对数据的初步分析以及专家经验,当两架航班相关,在时间轴上先行的航班出现延误时,则极易发生延误波及。设 A、B 为两架相关航班,A 与 B 的相关性主要分以下三类:

(1) 飞行器相关,即 B 航班与 A 航班使用的是同一架飞行器,航班号不同,而飞行器编号相同。简单讲就是同一架飞机,执行两次连续的飞行任务。

(2) 机组及服务人员相关,即 A 与 B 采用的是同一组飞行人员,或机舱乘务人员。

(3) 旅客相关,即 B 航班中包含有从 A 航班上转乘的旅客。

另外,此种相关具有传递性,即 A 与 B 相关,B 与 C 相关,则 A 与 C 相关。因此当一架航班发生延误,尤其是延误时间较长时,由于延误波及,可能会造成多架相关航班、在多个不同机场发生延误。

三种航班相关中,第 2、3 类相关可以尽量采取方法人为地避免,如利用航空公司的排班系统,而最为常见,也最难以避免的是第 1 类。故在本研究中,主要关注第 1 类相关,即飞行器相关。图 1 为延误波及现象示意图,某两次连续飞行任务,由同一架飞机执行,两次航班的进、离港延误,以及其间的延误波及情况如图 1 所示。

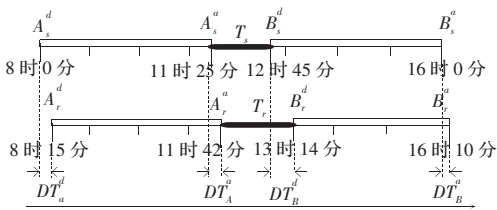


图 1 连续航班的延误波及示意图

2.2 延误波及模型分析

在航班计划中,B 航班是 A 航班的下游航班。A 的计划起飞时间为 A_s^a ,B 的计划起飞时间为 B_s^d ,当 A 航班的实际起飞时间 A_s^d 晚于计划起飞时间 A_s^a ,即 A 产生了离港延误,相应地 A 进港也发生了延误:

$$DT_A^d = A_r^d - A_s^d \quad (1)$$

$$DT_A^a = A_r^a - A_s^a, \quad (2)$$

其下游航班 B,由于与 A 相关,受到 A 航班进港延误的波及,其实际起飞时间为 B_s^d ,A 的进港延误波及到 B,使 B 产生了一系列离港、进港延误:

$$DT_B^d = B_r^d - B_s^d \quad (3)$$

$$DT_B^a = B_r^a - B_s^a \quad (4)$$

B_s^d 与 A 航班的进港时间 A_s^a 的差值称为航班的过站时间(或周转时间)。过站时间主要包括地面服务和冗余时间。则由定义可知:

$$T_s = B_s^d - A_s^a = t_{gs} + \Delta t \quad (5)$$

$$T_r = B_r^d - A_r^a = t'_{gs} + \Delta t' \quad (6)$$

其中 T_s 和 T_r 分别表示 A、B 航班间计划过站时间和实际过站时间。 t_{gs} 表示地面服务所需的时间, Δt 是预留的冗余时间。

由式(5)可推出:

$$B_r^d = A_r^a + T_r \quad (7)$$

将式(3)、式(4)代入式(7),可得:

$$DT_B^d = B_r^d - B_s^d = A_r^a + T_r - (A_s^a + T_s) = DT_A^a - (T_s - T_r) \quad (8)$$

这就是延误波及的数学表达式。

当延误波及发生时,通常 $T_r \neq T_s$,且:

$$\begin{cases} DT_B^d \leq DT_A^a, & T_r \leq T_s \\ DT_B^d > DT_A^a, & T_r > T_s \end{cases} \quad (9)$$

$$\begin{cases} DT_B^d \leq DT_A^a, & T_r \leq T_s \\ DT_B^d > DT_A^a, & T_r > T_s \end{cases} \quad (10)$$

在理论上,我们希望实际过站时间小于计划过站时间,见式(9),这样可以对延误进行吸收,以保证 B 航班因波及而产生的离港延误时间小于 A 航班的进港延误。但在实际情况下往往并非如此,参见图 1 中,此两次连续航班的周转时间长于计划周转时间。这是由于往往因跑道、地服等资源所限,为保证其他航班的正点起飞,而人为地使延误航班过站的冗余时间 Δt 延长,即当其延误的发生已无法避免时,其服务优先级会被降低。

2.3 航班延误主要因素

造成航班延误的因素众多,航班延误的因素按其属性可以分为四大类^[2]:航空公司原因、机场管理原因、航空管制原因、以及旅客原因,每一类下又分为若干小类,见表 1。而在航空公司的记录中,对延误原因的分类更加详细,达 20 余种之多。

表 1 航班始发地机场延误因素分类

	1. 机票超售
1. 航空公司管理	2. 运力调配
	3. 机械故障
2. 机场管理	1. 地面服务保障
	2. 机场计算机系统
3. 航空管制	1. 流量控制
	2. 天气预报
4. 旅客	1. 旅客晚到
	2. 旅客突发事件

航班延误本身的因素本已十分复杂,由于延误波及而产生的延误,其形成因素将更加复杂,由 1.2 节中的式子可得出:

$$DT_B^d = f(A_r^a, t'_{gs}, \Delta t', A_s^a, t_{gs}, \Delta t) \quad (11)$$

各因素之间仍有关联,如:

$$A_r^a = g(A_s^a, DT_A^d, d, v_A) \quad (12)$$

其中 d 为航程距离, v_A 为 A 航班的飞行速度, d 一定时, v_A 为有上下限的可变量,同一机型的上下限一致。

此外,成因中还包含有很多不确定因素,那么这些复杂的函数关系在数学上的量化和表述都是很困难的。于是选用了适用于解决不确定因果关系的贝叶斯网络来进行学习。贝叶斯网络常用于研究在复杂系统中多因素间的相互依赖,它具有在推论和可视化两方面的独一无二的强壮性。

3 基于贝叶斯网络的延误波及分析

3.1 贝叶斯网络 BN(Bayesian net)

贝叶斯网络(亦称贝叶斯置信网络 Belief Network、概率网络或者因果网络)是于 1985 年由 Judea Pearl 首先提出,是目

前在人工智能中一种很重要的推理技术。

贝叶斯网络的理论依据是概率论,其网络拓扑结构是一个有向无环图(DAG),结点代表随机变量,有向弧代表随机变量间的条件依赖关系。各变量间满足条件独立性的前提下,联合概率分布可以表示如下:

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i | Pa_i)$$

在研究所构建的贝叶斯网络中,由于需要多项式随机变量来对 θ 进行贝叶斯估计,则对 θ 的最大似然估计为多项似然函数,狄利克雷分布是多项似然函数共轭分布的,因此设 $p(\theta)$,即先验分布为狄利克雷 $dirichlet(\alpha_1, \alpha_2, \dots, \alpha_p)$,其中 α_i 为狄氏分布的超参数, p 为状态数。又因为它是共轭的,它的后验分布仍符合狄氏分布,故对随机变量假设特定的父值,观察 i 个状态后,其后验分布则为: $dirichlet(\alpha_1 + n_1, \dots, \alpha_p + n_p)$, n_i 为充分统计量。由先验分布将模型中每个节点的所有可能状态,都设置为相等的狄氏密度(可设 $\alpha_1 = \alpha_2 = \dots = \alpha_p = 1$)。这个方法适合于变量为绝对随机的、且不同类别的参数都无关时使用。获取的很多参数是连续变量,如所有的时间变量,需要先行离散化后在使用。

另外,由于在航空公司的数据库中,总会有因各种原因而没有观测或录入的数据,并且这种缺失都是随机的,所以所获得的数据永远是不完全的。于是采用 EM 算法(期望最大化算法)。EM 算法的根本思想是用当前的参数估计值 θ^l 来修补当前不完整的数据 D ,使其完整。修补后有几种可能取值,给每个修补后的可能数据一个权值 $w_i = P(X_i = x_i | D_i, X_j = x_j)$, X_i 是 D_i 中所有缺值变量的集合, x_i 是 X_i 的一个取值。根据修补后的数据,计算其出新的最大似然估计 θ^{l+1} 。在算法的实现方面,EM 算法属于一种迭代算法^[6],当迭代满足收敛条件后,将返回 θ^{l+1} 的值。

3.2 数据的选取与前期处理

目前繁忙与非繁忙机场的延误情况差距很大,北京、上海虹桥、成都、深圳、大连、乌鲁木齐等 18 个机场由于地面及空域资源紧张,运行已经处于饱和状态,特别是首都机场,高峰小时内不到一分钟就有一个航班起降。延误波及的源头大都发生在近乎饱和或已经超负荷运行的机场中,且一旦有延误波及现象发生时,对这些繁忙机场的影响非常严重。而非繁忙的机场航班延误现象相对较少,部分延误的出现也是受到繁忙机场延误的波及所致。有鉴于此,减少繁忙机场的延误波及的影响,就可以使整个航空系统的航班延误得到缓解,所以研究对象主要集中在繁忙的国际枢纽型机场。

本次研究所采用的数据来源于国内某大型航空公司 2006 年 3 月 26 日至 4 月 7 日的所有航班记录,以北京机场为主要研究对象,从中筛选出符合条件的记录,并通过离散化等前期数据处理。

根据相关规定,北京机场的进、离港航班,延误时间在 0 至 20 分钟的,视为正常航班,其他都作为非正常航班。而在非正常航班中,有少部分航班是属于提前到达,那么他们的非正常是不会对下游航班的正常离港造成任何影响,即不会产生延误波及。因此将航班的提前到达用负数表示(如提前到达 10 分钟,即延误为 -10),那么所有小于 20 分钟、包括负数的进、离港航班,在研究中均视为正常。

在数据的前期处理过程中,可直观地发现凌晨、深夜中航架架次较少,相对地,延误发生的也较少。则离散后的时间变量

划分成的三个状态:0 点至 8 点、8 点至 21 点、21 点至 23 点。将延误时间分为 4 个等级:<20 分钟为正常;20 分钟~40 分钟为轻度延误;40 分钟~60 分钟为中度延误;≥60 分钟为重度延误。

另外,在前期处理中还发现,当延误严重和发生大面积延误的时候,常见到“航班取消”的记录。

3.3 进港延误(Arrival Delay)对离港延误(Departure Delay)的波及

进港延误的主要影响因素可由上述的式(12)分析得出,因空中交通的特点,航线距离 d 相对固定,航行速度 v 只在一个可知的小范围(与机型有关)内变化,所以航班在空中的航行时间相对固定。简单地说是离港的正点得到保证了,才可以保证进港的正点。故我国各大航空公司与机场普遍比较关注的是离港延误,所以将离港延误作为研究的焦点,在贝叶斯网络模型将其设为因变量。

通过 2.2 节中的分析,由式(8)可知, $DT_B^d = DT_A^d - (T_s - T_r)$,在飞行器相关的连续航班中,A 的进港航班延误对 B 的离港延误影响巨大。当进港有轻度延误时,即等式右边第一项值较小,可以通过一系列缩短过站的手段,即增大第二项,将轻度延误吸收、甚至消除,使延误不再波及或波及较小。但若进港为中、重度延误,即第一项数值较大时,而第二项的值不能无限增大: $\max(T_s - T_r) = T_s - \min T_r < T_s$,其中 T_s 为定值,则波及基本无法避免,离港必会因为进港延误的波及而产生延误。特别当第二项为负值时,波及所产生的延误将放大。

根据专家经验来构建贝叶斯网络的结构^[7],并使用 Netica^[8]建立贝叶斯的网络拓扑如图 2 所示。通过贝叶斯网络的学习,发现 93.9%的进、离港延误都发生在一天中的 8 点~21 点之间,只有 2.8%发生在凌晨 0 点~8 点,3.3%发生在晚上 21 点~23 点。两种状态的概率之和仅占 6.1%,所以需要研究的主要对象是 8 点~21 点间的延误波及情况。

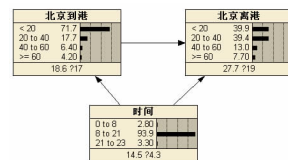


图 2 进、离港拓扑图

如图 3,8 点~21 点间若进港无延误,则离港航班正常无延误的概率为 47.3%,轻度延误的为 42.4%。若进港为轻度延误,如图 4,则离港延误也主要集中在轻度延误,为 37.3%,其次为中度延误(40 分钟~60 分钟),约为 30%,仍有 24%的航班,在进港轻度延误的情况下,可以正常离港。

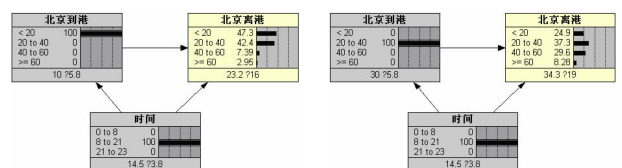


图 3 进港正常

图 4 进港轻度延误

当进港延误为中度时(40 分钟~60 分钟),见图 5,则离港延误主要集中在轻度、中度、重度延误,分别为 25.9%、36.2%、25.9%。当进港延误为重度时(≥60 分钟),则离港延误也是重度的概率为 62.9%,见图 6。

可以分析出在先行航班的进港延误程度不同时,延误波及的影响程度不同。

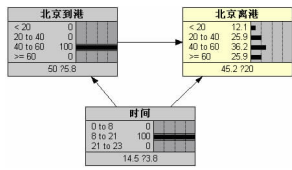


图5 进港中度延误

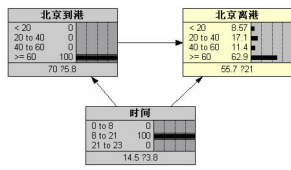


图6 进港重度延误

3.4 航班取消对离港延误的波及

当延误严重时,常常采用的措施是航班取消。当然航班取消本身可能还有其他的成因,如经济因素等,这不在研究的范畴之内。

将航班取消作为影响本地离港延误的重要因素,引入BN网络拓扑中来,通过学习可发现,当无延误时,航班的取消与否不影响延误的波及,对比图7(a)、(b)。

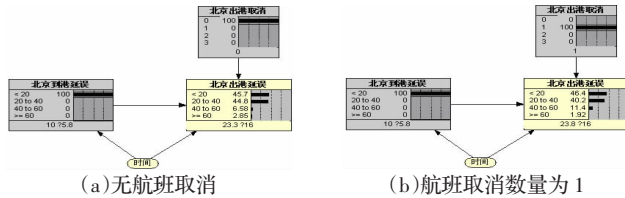


图7 进港为正常

当轻度延误时,航班取消会使波及程度有所减弱,但不明显,对比图8(a)、(b)。

当进港延误为中度时,航班取消减轻了波及的影响,使离港延误有所改善,对比图9(a)、(b)。

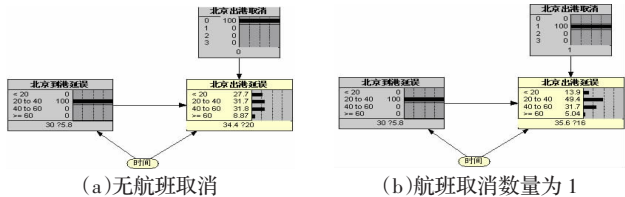


图8 进港为轻度延误

重度延误的波及受到航班取消的影响,可使波及减弱,对比图10(a)、(b)。

取消2架航班的影响大于取消1架航班。可见多架航班的取消可在重度进港延误情况下,明显减少对离港延误的影响。

取消航班的确能在不同的延误波及情况下减少延误波及的影响。并且可基于本研究,为取消航班及取消架次的决策提供参考。

4 结束语

论文以延误问题较严重的某国际机场为研究对象,初步研

(上接 211 页)

[2] 滕树杰,张乃尧,崔震华.核动力装置蒸汽发生器水位的分层模糊自适应控制[J].控制与决策,2002,17(6):933-936.

[3] 郭暹凤.液位压力连锁自动排水控制系统的改进[J].西部探矿工程,2001(增刊):338-340.

[4] Astrom K J,Ostberg A B.laboratory for process control[J].IEEE Control Systems Magazine,1986,6:37-42.

[5] Malmberg J,Eker J.Hybrid control of a double tank system[C]// Proceedings of International Conference on Control Applications, Hartford, CT, 1997: 133-138.

[6] 袁浩,段瑞峰,闫淑英.一种新型的数字式液位-流量控制系统综合实验装置[J].实验技术与管理,2002,19(4).

[7] 黄慎之,顾训,胡睿.一种用于实验室的液位过程控制系统[J].上海大学学报:自然科学版,2001,7(5):412-416.

[8] 李洪兴,王加银,苗志宏.模糊控制系统的建模[J].中国科学:A辑,

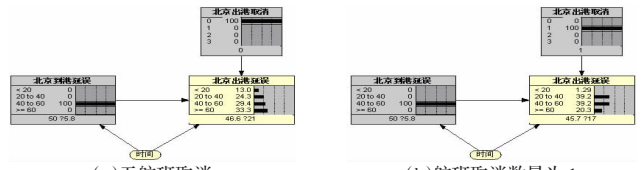


图9 进港为中度延误

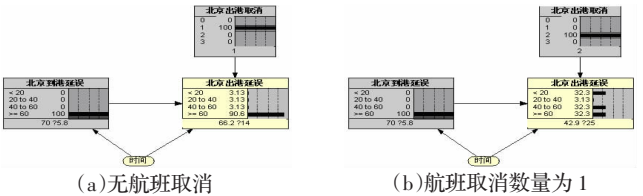


图10 进港为重度延误

究了延误波及对航班延误的影响。基于贝叶斯网络的学习反映出延误波及的存在,以及在各种不同状态下,延误波及对下游相关航班的影响程度。并能够通过该模型,在一定程度上预测出波及后造成的延误程度。

本次获得的数据集中在3、4月份,此阶段的气候比较稳定,气候因素对延误以及延误波及的影响不大。未来在获取更多数据的前提下,需要将天气等其他加入到网络拓扑之中。

参考文献:

[1] Xu Ning,Chen C H.Propagation of Delays in the National Airspace System[C]//the 22nd Conference on Uncertainty in Artificial Intelligence,2006-07.

[2] 石丽娜.多等级模糊评价方法在航班延误中的应用[J].上海工程技术大学学报,2006,20(3):276-279.

[3] Schaefer L,Millner D.Flight delay Propagation analysis with the detailed policy [C]//the 2001 IEEE Systems,Man,and Cybernetics Conference,2001.

[4] Xu Ning,Donohue G,Laskey K B.Estimation of delay propagation in the national aviation system using Bayesian networks [C]//6th USA/Europe Air Traffic Management Research and Development Seminar,2005-06.

[5] Netica[EB/OL].http://www.norsys.com/.

[6] 张少中,章锦文,张志勇,等.面向大规模数据集的贝叶斯网络参数学习算法[J].计算机应用,2006,26(7):1689-1692.

[7] 张克,汪云甲.基于贝叶斯网络的煤与瓦斯突出预测研究[J].计算机工程与应用,2007,43(29):220-221.

2002,32(9):772-781.

[9] 苗志宏,李洪兴.一类模糊系统的逼近问题[J].北京师范大学学报:自然科学版,2000,36(1):14-21.

[10] 李洪兴.模糊控制的插值机理[J].中国科学:E辑,1998,28(3):259-2673.

[11] 李洪兴,王加银,苗志宏.模糊控制系统的建模中的边缘线性化方法[J].自然科学进展,2003,13(5):466-472.

[12] 金梅,关学中,金中石.一类非线性系统的模糊建模与仿真[J].系统仿真学报,2005,17(5):1030-1047.

[13] 李洪兴.从模糊控制的数学本质看模糊逻辑的成功[J].模糊系统与数学.1995,9(4):1-14.

[14] 王志新,谷云东,王加银,等.双容水箱上的几种液位控制实验及被控对象的数学模型[J].北京师范大学学报:自然科学版,2006,42(4):5-10.