Contents lists available at ScienceDirect

Engineering

journal homepage: www.elsevier.com/locate/eng

Research
Novel Methodologies in Air Transportation—Article

基于张量因子分解框架的航班延误模式分析

张明远[a,b]，陈莘文[a,b]，孙立君[c]，杜文博[a,b,*]，曹先彬[a,b,*]

[a] School of Electronic and Information Engineering, Beihang University, Beijing 100191, China
[b] National Engineering Laboratory for Comprehensive Transportation Big Data Application Technology, Beijing 100191, China
[c] Department of Civil Engineering and Applied Mechanics, McGill University, Montreal, QC H3A 0C3, Canada

ARTICLE INFO

Article history:
Received 10 April 2020
Revised 21 July 2020
Accepted 3 August 2020
Available online 19 March 2021

关键词
空中交通管理
航班延误
潜在类别模型
张量分解

摘要

在空中交通管理和民航机场运营过程中，从过去的运营中获得的经验对于设计合理的策略至关重要。因此，本文使用大量时空飞行数据来识别相似的航空交通和延误模式，这对于更好地了解航空系统和制订管理决策至关重要。但是，由于数据集隐含了复杂的依赖关系以及高维时空之间的交互作用，因此挖掘数据中的重要特征和模式非常具有挑战性。本文提出了高维历史飞行数据的张量因子分解框架。我们通过潜在类别分析模型，并使用2014—2017年中国224个机场的空中交通运行数据证明了该框架的有效性。研究发现每个维度可以清晰地代表不同的运行模式。为了证明这些模式的有效性，我们创建了一个估计模型，该模型可实现对航班延误程度的初步判断。上述结果可以根据历史情景获得的经验，帮助机场运营商和空中交通管理人员更好地了解空中交通和延误模式。

1. 引言

随着世界民用航空业的飞速发展，严重的航班延误仍然是一个重要问题。航班延误不仅使乘客不愿考虑航空交通或再次选择同一家航空公司[1–4]，而且迫使航空公司承担飞机维护和机队利用率不足的额外费用[5]。此外，航班延误导致燃油消耗和二氧化碳排放量增加，对环境造成了极大危害[6,7]。除了上面列出的直接影响外，航班延误对整个社会经济发展都具有负面影响[8]。

许多因素使得此问题变得更复杂和棘手，这些因素通常为异常天气[9]和技术原因[10]，其中技术原因主要包括空中交通管制[11]、设施容量不足或者调度不当[12]、运行程序更改[13]和缓冲时间不足[14]等。因此，挖掘潜在的航班延误模式并设计适当的策略非常困难[15]。最近，基于历史观测数据的分析方法已被证明不受上述约束影响并适用于包含隐藏信息的动态数据[16]。因此，促进系统认知和决策的方法是充分利用历史数据[17]。例如，当遇到恶劣天气时，可以查询天气条件相似的过去几天，并参考当日空中交通管制员采取的行动。近期的几项研究[18–23]致力于发现空中交通管理中的固定模式。Liu等[19]介绍了一种半监督学习算法，可以将相似的日期划分为不同的模式。第一步先量化每小时航空天气预报数值之间的距离（表示相似性），然后确定总距离较小的日期。他们应用此方法在纽瓦克自由国际机场（EWR）进行了两个案例研究，并证明了其有效性。Mukherjee等[20]也提出了一种根

* Corresponding authors.
E-mail addresses: wenbodu@buaa.edu.cn (W. Du), xbcao@buaa.edu.cn (X. Cao).

英文原文: Engineering 2021, 7(4): 465–472
引用本文: Mingyuan Zhang, Shenwen Chen, Lijun Sun, Wenbo Du, Xianbin Cao. Characterizing flight delay profiles with a tensor factorization framework. *Engineering*, https://doi.org/10.1016/j.eng.2020.08.024

据航空天气条件对运行模式进行分类的方法。他们使用天气指数作为输入并应用因子分析来确定主要的天气模式，然后，使用Ward的最小方差法将日期聚类。属于同一模式的日期共享相似的天气模式。除天气条件外，一些研究还试图从其他角度确定相似的日期。Grabbe等[21]使用*k*-means聚类算法来识别地面延误程序中的相似日期，并对地面延误程序的开始和结束时间以及计划的到达率数据应用期望最大化（EM）算法。其他类似的研究主要关注空中交通流量和航班延误以确定相似的模式[24]。Gorripaty等[17]测量了需求和容量数据中的主要成分，但对数据进行聚类分析后发现，需求或容量中没有固定的模式。通过确定国内直飞航班到达延误的周期性模式，Abdel-Aty等[25]发现统计方法未能有效检测某些模式。

尽管做了一些研究，但是这些研究在理解航班延误模式方面仍然存在差距。如上所述，一种有效的方法是在时空历史数据中找到聚类模式[26,27]。但是，由于该数据的高维特性，其很难在欧式空间中找到明显不同的模式[28]。因此，后续研究提出了潜在成分分析的方法来揭示隐藏的模式，诸如潜在分布分析[29]、潜在特征分析（包括响应理论和Rasch模型）以及层次分析等方法，可以利用从张量分解中获得的特征来形成投影和维度较低的子空间，以增强时空交通动态模式的底层聚类结构[30]。这些方法开辟了交通科学领域的新方向[31]，如城市流动性分析[32–34]、交通速度预测[35]、交通数据缺失[36,37]和船舶航迹恢复[38]。

受上述方法的启发，本文的主要研究目标是使用大量的空中交通数据来了解潜在的空中交通和航班延误模式。首先将飞行记录数据视为从泛分布中抽取的多元观测概率数据，将概率分解问题等价为Tucker潜在类别分析模型来挖掘主要模式。然后，提出一个估计模型用于在已知信息极少的情况下判定延误程度。本文的其余部分安排如下：第2节介绍模型框架；第3节展示一个基于中国航空数据的案例研究；第4节给出主要结论。

2. 模型框架

本节介绍了以概率为基础的飞行数据建模的总体框架。目的是从不同角度描述空中交通和航班延误的主要模式及其相互作用。第2.1节介绍了在框架中使用的数据表示方法；第2.2节提出了一种非负Tucker分解（NTD）方法；第2.3节描述了潜在类别分析（LCA）方法。

2.1. 数据表示

我们令$x_a = (x_{a1}, ..., x_{aq})^T$代表一个飞行记录a，其中$q$代表维数，即单个行程记录中的属性维度。为了表征航班的特征，每个元素可以表示航班的离港机场（x_{a1}）、离港时间（x_{a2}）、离港日期（x_{a3}）、延误程度（x_{a4}）等。

为方便起见，将这些值映射为离散值。$x_{a\beta} \in \{1, ..., w_q\}$为属性$\beta$（$\beta = 1, ..., q$）从1开始的离散值。$\beta$表示属性的索引。$w_q$表征维度为$q$的离散值向量。以离港机场为例，对于记录a来说，$x_{a1} = 1$代表机场1。$x_{a2} \in \{1, ..., 24\}$表示航班的离港时间，每个值对应一天中的一个小时。然后，使用4年（2014—2017年）的飞行记录数据介绍我们的方法，其中所有航班的离港日期按时间顺序编号为$x_{a3} \in \{1, ..., 1461\}$。由于到港延误与空中航班时段密切关联，可能会受到空中飞行加速、减速的影响，因此本文考虑采用离港延误而非到港延误来研究航班延误模式，以更好地反映目标机场的运行状态。根据美国联邦航空管理局（FAA）制定的规则及相关研究，延误航班是指比计划时间晚15 min以上起飞的航班。考虑实际情况，本文选择较计划离港时间晚45 min和90 min作为阈值。因此，本文将每个航班的离港延误分为4个级别：① < 15 min；② 15~45 min；③ 45~90 min；④ > 90 min。我们分别用$x_{a4} \in \{1, ..., 4\}$表示“准时”“轻度延误”“中度延误”和“重度延误”。

2.2. 非负 Tucker 分解

已有研究表明，张量分解在各种情况下都显示出许多优点，尤其是当必须将数据分解为加性成分之和时[39]。张量分解最早由Tucker [40]在1963年提出，非负NTD [41,42]是在张量分解基础上获得的，用于处理自然数据的非负观测值。NTD是一个强大的工具，可以从高维张量数据中提取基于非负数部分的潜在分量，同时保留数据的多线性结构[24]。从数学维度来说，它将张量分解为一组矩阵和一个核心张量。

给定K阶张量χ，NTD将任意一个非负K阶张量$\chi \in R_+^{I_1 \times I_2 \times \cdots \times I_K}$（$R_+$是正实数空间；$K$是空间维度；$I$是正交基）分解为非负核心张量$\varsigma \in R_+^{J_1 \times J_2 \times \cdots \times J_K}$（$J$是正交基）和$K$个非负矩阵的模积$A^{(K)} \in R_+^{I_K \times J_K}$：

$$\begin{aligned} \chi \in \varsigma \times {}_1A^{(1)} \times {}_2A^{(2)} \times {}_3A^{(3)} \times \cdots \times {}_KA^{(K)} \\ = [\varsigma, A^{(1)}, A^{(2)}, \ldots, A^{(K)}] \end{aligned} \quad (1)$$

式中，$A^{(1)}$, $A^{(2)}$, $A^{(3)}$, ..., $A^{(K)}$被称为因子矩阵；ς是核心张量，显示了不同因子矩阵之间的交互作用和连接程度。在这种方

法中，核心张量ς和因子矩阵$A^{(K)}$在元素上为非负数。具体如下：

$$\chi_{\eta_1,\eta_2,\ldots,\eta_K}=\sum_{r_1=1}^{R_1}\sum_{r_2=2}^{R_2}\cdots\sum_{r_K=1}^{R_K}\varsigma_{r_1r_2\cdots r_K}\partial^{(1)}_{\eta_1r_1}\partial^{(2)}_{\eta_2r_2}\cdots\partial^{(K)}_{\eta_Kr_K}\quad(2)$$

式中，$\partial^{(K)}_{\eta_Kr_K}$是一个因子矩阵；$R_K$为因子矩阵的大小；$r$和$\eta$是因子矩阵中模式$K$的维数。将分解建模为优化问题：

$$\begin{aligned}&\text{Minimize }\frac{1}{2}||\varsigma\times{}_1A^{(1)}\times{}_2A^{(2)}\times{}_3A^{(3)}\times\cdots\\&\qquad\times{}_KA^{(K)}-\chi||_F\\&\text{subject to }\varsigma\geq0,\ A^{(K)}\geq0\end{aligned}\quad(3)$$

式中，F为矩阵范数。

2.3. 潜在类别分析

潜在类别分析是一种统计方法，可从多元分类数据中找到潜在类别[43,44]。潜在类别模型公式如下：

$$P_{i_1,i_2,\ldots,i_N}=\sum_{t}^{T}p_t\prod_{n}^{N}p^t_{\partial\mu}\quad(4)$$

式中，$P_{i_1,i_2,\ldots,i_N}$表示概率分布方程；i是维度指标；T是每个维度类别的模式数量；N是维度数量；t是每个维度模式的索引；p_t是加为1的概率；$p^t_{\partial\mu}$是条件概率；∂和μ是概率矩阵的维度。潜在类别分析通过条件独立性的标准来定义潜在类别。这意味着每个变量在统计上独立于每个潜在类别中的每个其他变量。因此，可以将概率张量中的每个元素计算为所有模式组合的总和。

$$P_{i_1,i_2,\ldots,i_N}=\sum_{t}^{T}p_t\ \theta^{(1)}_{i_1t}\times\cdots\times\theta^{(N)}_{i_Nt}\quad(5)$$

式中，$\boldsymbol{\theta}^{(N)}$是表示模式$N$的概率向量。

在本研究中，我们使用了潜在类别模型，该模型假设每个观察值都是由基础类别的混合生成的，并且每个类别都与唯一的概率分布相关。因此，联合分布被认为是乘积多项式与观测概率的混合。使用第2.1节中的符号，可以将所有飞行记录x汇总为一个维度$\phi=w_1\times w_2\times\ldots\times w_m$的$m$阶张量，并且张量中的每个单元格$(v_1, v_2, \ldots, v_m)$（$v$代表张量的一个维度）都是对飞行数量$\sum\delta(x_{a1}=v_1,\ldots,x_{am}=v_m)$的计数。$\delta$是一个二值的指示函数，$\delta=1$为真，$\delta=0$为假。为了更好地理解数据集的内部联系，我们将这些飞行记录放入一个概率张量中，其每个值代表一个飞行记录属于该坐标的概率。每个值的概率张量表示属于特定单元格的飞行概率$p_c(x_{a1}=v_1,\ldots,x_{am}=v_m)$。观测概率（也是概率质量函数）可以通过Tucker分解以类似的方式重新生成。

$$P(x_a|\theta)=\sum_{r_1=1}^{R_1}\sum_{r_2=1}^{R_2}\cdots\sum_{r_N=1}^{R_N}\pi_{r_1r_2\cdots r_N}\prod_{n=1}^{N}\boldsymbol{\theta}^{(N)}_{x_{an}r_n}\quad(6)$$

核心张量π捕捉了不同维度模式之间的交互性。$\boldsymbol{\theta}^{(N)}_{x_{an}r_n}$是概率因子矩阵，代表了维度$N$的主要模式。可以将概率张量中的每个元素计算为所有模式组合的总和。

$$\begin{aligned}&P(x_{a_1}=v_1,\ldots,x_{a_m}=v_m|\boldsymbol{\theta})\\&\quad=\sum_{r_1=1}^{R_1}\sum_{r_2=1}^{R_2}\cdots\sum_{r_N=1}^{R_N}\pi_{r_1r_2\cdots r_N}\boldsymbol{\theta}^{(1)}_{i_1r_1}\boldsymbol{\theta}^{(2)}_{i_2r_2}\cdots\boldsymbol{\theta}^{(N)}_{i_nr_n}\end{aligned}\quad(7)$$

$\boldsymbol{\theta}^{(N)}_{i_nr_n}$是表征维度为$n$的模式分布的概率矩阵。该模型等效于非负Tucker（NNT）分解[45,46]，可以识别不同维度的通用模式，并通过核心张量揭示相互作用。EM算法可以有效推导该模型[47]。

3. 案例研究

3.1. 航班运行数据

本文分析的数据集由中国民用航空局（CAAC）提供。鉴于本文的研究目的是为空中交通和机场管理提供决策支持，因此航班延误情况是本文研究重点。而且，空中交通流量是延误情况的基础。因此，本文选择空中交通流量和航班延误作为主要研究对象。选择离港延误是因为其可以更好地反映离港机场和空域的拥挤程度。表1列出了航班数据的分类值。该数据库包含13 492 326架国内航班。所有航班共连接224个机场，其中北京首都国际机场的航班数最多，占所有机场航班数的6.3%。航班的出发日期为2014年1月1日—2017年12月31日，出发时间可以为一天之内的任何时间。可以看出，根据出发延误时间可以将所有航班分为4组。4个组中的航班数占比分别为37%、38%、14%和11%。这些航班的平均离港延误时间为31.08 min。2008年2月9日是航班数量最多的一天，为12 419架次。2014年1月1日是航班数量最少的一天，为7009架次。机场最为繁忙的时间段是8:00~9:00，其中有6%的航班在此时间段内起飞。

3.2. 张量分解

本文假设飞行记录是从通用分布中采样的多元变

量。将13 492 326次飞行记录汇总为同一张量。每个观测值包含4个变量，包括离港机场、离港日期、离港时间和延误程度。总组合为224 × 1461 × 24 × 4。在这里，我们使用一个大小为3（离港机场，A）× 4（离港日期的编号，D）× 5（一天中的时间，H）× 4（延误程度，L）的核心张量π来捕获不同模式间的相互作用。尽管更大的核心张量可以包含更多信息并反映不同模式之间的全面关系，但是较小的核心张量可以促进对结果的解释。此外，现有研究表明，结果在不同尺寸的核心张量上基本一致[48]。在下文中，我们将以核心张量大小[3 × 4 × 5 × 4]为例介绍主要结果。

图1（a）描绘了5个模式的离港时间分布。模式H_1占所有航班的18.7%，其从11:00开始逐渐上升，并在24:00达到峰值。模式H_2与高斯分布的形状相似，其在17:00达到峰值，并且该模式占据所有航班的24.4%。与H_1和H_2相比，H_3和H_5分布更加集中。它们在早上显著增加，到中午突然下降。模式H_4在11:00达到峰值，然后在一天的其余时间连续下降。H_3、H_4和H_5的比例分别为22.9%、22.2%和11.8%。离港日期的因子矩阵由1461行和4列组成，这些因子矩阵描述了原始离港日期和离港日期模式之间的对应关系。我们分析了一年中不同月份和一周中不同日期的模式，并汇总了一年中不同月份和一周7天的日期模式的分布概率，以确定流量分布如何与离港日期交互。一周7天的日期模式如图1（b）所示。模式W_1集中在工作日，而模式W_3的峰值出现在周末两天。模式W_2和W_4则呈现相反趋势。W_2主要集中在星期一、星期六和星期日，模式W_4主要集中在星期三、星期四和星期五。图1（c）显示了不同月份的模式。我们可以观察到明显的季节多样性。模式M_1主要集中在秋冬季，而模式M_2主要集中在冬季和春季。模式M_3集中在春季，而模式M_4主要分布在夏季，夏季是航空系统的旅行高峰时段。

虽然上述计算过程未考虑任何空间位置信息，我们仍可以通过机场的地理位置来识别这些模式的特点。作为中国的交通枢纽，北京首都国际机场以A_1和A_2模式为主。此外，模式A_1主要分布在东南部地区，而模式A_2主要分布在西南部地区。模式A_3主要由中西部地区的机场组成。

图2显示了延误程度因子矩阵中每个模式（列）的延误程度的组成。L_1、L_2、L_3和L_4代表不同延误程度的航班，从“准时”到“重度延误”（如2.1节所述）不等。与原始张量相比，延误程度模式几乎保持不变，但存在不同延误程度的航班。图2显示了每个延误程度模式的组成，分别占总样本数量的40.7%、10.5%、29.1%和19.7%。应当指出，延误是空中交通拥堵的一种表现，因此延误程度与时空因素中的交通流量特性密切相关。

表1 飞行数据的分类值

Attribute	Category	Description
Departure airport	224	Origin airport
Order of day	1461	Ordered by date during four years
Time of day (h)	24	0:00–1:00, 1:00–2:00, …, 23:00–24:00
Level of delay	4	① < 15 min, ② 15–45 min, ③ 45–90 min, ④ > 90 min

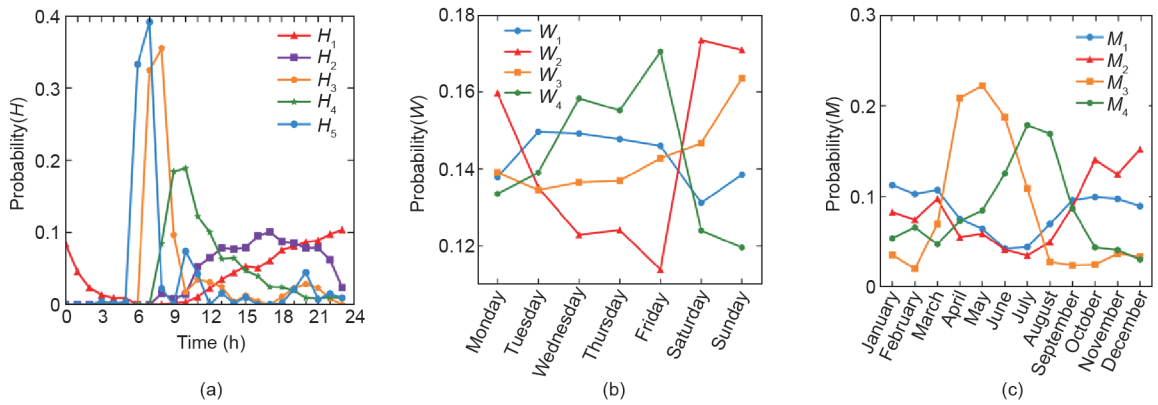

图1. 不同维度的主要模式。（a）每个模式（列）的离港时间在因子矩阵中的概率分布Probability(H)；（b）离港日期在星期几的概率分布Probability(W)；（c）离港日期在月份的概率分布Probability(M)。

为了进一步分析，我们采用延误程度模式和其他模式之间的条件概率来研究相互作用。我们根据贝叶斯定理计算条件概率分布Probability($L|H$)。给定离港时间模式，延误程度模式的条件分布为：

$$\text{Probability}(L|H)=\begin{bmatrix} & L_1 & L_2 & L_3 & L_4 \\ H_1 & 0.2857 & 0.4294 & 0.0997 & 0.1852 \\ H_2 & 0.3894 & 0.0340 & 0.2771 & 0.2990 \\ H_3 & 0.4049 & 0.0000 & 0.4270 & 0.1681 \\ H_4 & 0.4017 & 0.0585 & 0.3197 & 0.2201 \\ H_5 & 0.6538 & 0.0290 & 0.2977 & 0.0196 \end{bmatrix} \quad (8)$$

Probability($L|H$)显示了延误程度模式如何与离港时间模式交互。可以看出，模式L_1表示“准时”时间与所有模式密切相关，这表明“准时”航班在一天中的任何时候都占据主流。H_1表示全天呈上升趋势的飞行流量模式，与延误模式L_2相关联，后者对应“轻度延误”的航班。H_2主要由离港时间模式L_3和L_4覆盖，表示高等级延误与下午的流量高度相关。H_3、H_4和H_5主要由L_3和L_4覆盖。这可以用以下事实来解释：由于大多数机场的空中交通繁忙，早晨离港高峰期间的航班可能会“重度延误”。

$$\text{Probability}(L|M)=\text{Probability}(L|W)=\begin{bmatrix} & L_1 & L_2 & L_3 & L_4 \\ M_1/W_1 & 0.3783 & 0.3433 & 0.2638 & 0.0146 \\ M_2/W_2 & 0.5889 & 0.3195 & 0.0780 & 0.0135 \\ M_3/W_3 & 0.4584 & 0.2777 & 0.1735 & 0.0904 \\ M_4/W_4 & 0.2223 & 0.2035 & 0.2560 & 0.3181 \end{bmatrix} \quad (9)$$

根据离港时间模式和延误程度模式的因子矩阵，我们发现在M_1/W_1、M_2/W_2和M_3/W_3中，“准时”和“轻度延误”模式最多。但是，M_4/W_4表示流量主要集中在工作日和夏季，并且往往会出现“重度延误”。

$$\text{Probability}(L|A)=\begin{bmatrix} & L_1 & L_2 & L_3 & L_4 \\ A_1 & 0.2369 & 0.2015 & 0.2830 & 0.2783 \\ A_2 & 0.3081 & 0.0417 & 0.4394 & 0.2107 \\ A_3 & 0.6402 & 0.0950 & 0.1434 & 0.1214 \end{bmatrix} \quad (10)$$

如式（10）所示，机场模式也与延误程度模式相关。显而易见的是，模式A_3主要由L_1覆盖，这可能表示中西部地区机场的空中交通延误较少。与A_3相比，A_1和A_2更可能产生“重度延误”，这可以用中西部地区机场相对较少的交通流量和充足的空域资源来解释。如上面的分析所示，延误受时间和空间的影响很大。为了进一步探讨模式之间的相互作用，我们提出了沿时间和空间维度的模式关联。如图3所示，(L_2, H_1)的值大于A_1中其他单元格的值，这表明当下午或傍晚的航班从东南部地区的机场起飞时，它们会稍有延误。我们还在(L_3, H_3)的A_2模式内观察到大量飞行流量，而这种流量很少出现在A_1和A_3中。由此可知，在上午高峰期，从西南部地区机场起飞的航班通常会有“中度延误”。如前所述，由于中

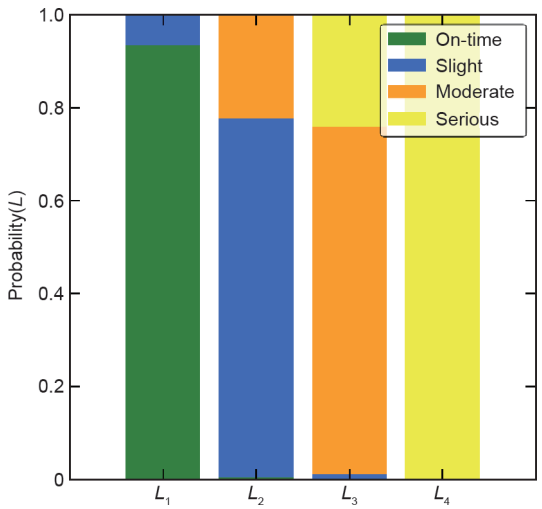

图2. 延误因子矩阵中每个模式（列）的延误程度概率Probability(L)。

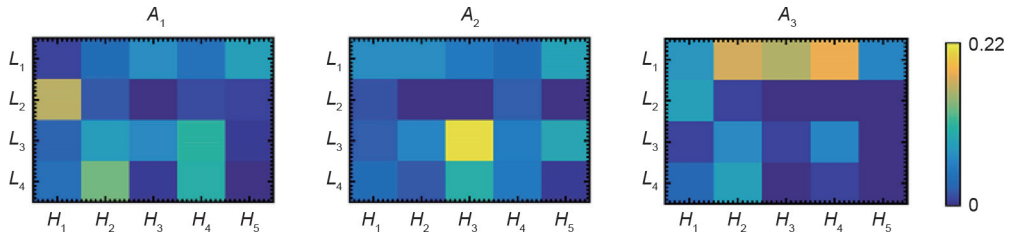

图3. 不同机场模式的延误程度和离港时间的关联。

西部地区机场（A_3）的运力过剩，航班延误的可能性很小，此现象与A_3中的单元格(L_1, H_2~H_4)一致。

实际数据表明，张量分解能够使我们解释基于潜在因素的复杂依赖性和高阶相互作用。核心张量π以非常有效且信息丰富的方式描述了不同模式之间的交互。该框架有助于我们理解和解释大型数据集中模式之间的潜在相互作用和复杂依赖性，从而加深我们对空中交通管理的理解。

在此基础上，存在另一个重大问题。如果仅考虑从时间和空间信息中提取的有关潜在模式的信息，是否可以估计延误程度？尽管以前的研究表明，航班延误可归因于许多复杂因素[49]，但这一问题可能具有重大意义。首先，由于各种延误原因的共同作用，产生了基于历史信息的潜在模式。例如，夏季极端天气频繁发生，并且夏季与严重的延误模式有强烈的相互作用。其次，由于时空信息不要求我们详细了解运行特性，因此，仅离港时间和离港机场等基本信息就可以帮助我们事先对延误进行初步评估。第三，航空系统的高度动态性和复杂性使人们相信，无法通过基本信息来估计延误。因此，如果实现了准确的估计，那么我们的框架所产生的潜在模式的有效性也得到了证明。由于该框架对这个问题的适应性，为了进一步探索，本研究采用随机森林（RF）算法来构建估计模型[50]。RF的优点包括：①可以对泛化误差产生内部无偏估计；②可在大型数据库上高效运行；③具有对变量之间相互作用进行建模的能力[51]。具体来说，RF是B颗树的集合$\{\Gamma_1(\boldsymbol{X}), ..., \Gamma_B(\boldsymbol{X})\}$，其中$\boldsymbol{X}=(x_1, ..., x_\rho)$是描述符的$\rho$维向量。集合产生$B$个输出$\hat{y}_1(x)=\Gamma_1(\boldsymbol{X}), ..., \hat{y}_B(x)=\Gamma_B(\boldsymbol{X})$，其中$\hat{y}_b(x)(b=1, ..., B)$是第$b$颗树的估计。汇总所有树的输出以产生一个最终估计$\hat{y}_{rf}^B(x)$。对于分类问题，$\hat{y}_{rf}^B(x)$是大多数树估计的类别。模型训练程序如下：

（1）准备训练数据。准备训练数据的ρ维样本及其类别标签。

（2）选择参数。E：每棵树的最大深度；C：分割内部节点所需的最小样本数；V：每次拆分的变量；M_L：在叶节点处需要的最小样本数。

（3）增长分类树。对于$b=1$~B，从训练数据中绘制一定尺寸（S_d）的引导样本Z^*。使用大约2/3的原始训练样本生长分类树，将剩下的1/3样本保留为所谓的袋外（OOB）样本。

（4）树的生成。对于每个引导样本，进行下述过程以生成一棵树$\Gamma_b(\boldsymbol{X})$。在每个节点上，选择最佳变量/分割点，并将该节点拆分为两个节点，直到该节点的样本数小于C，该树将增长到最大尺寸E，而不会被修剪。重复上述步骤，直到B颗树长出。

（5）结果输出。通过将树的估算值与森林中所有类似树的多数投票进行汇总来估算新数据。输出$\hat{y}_{rf}^B(x)=\text{majority vote}\{y_{b=1}^B(x)\}_{b=1}^B$，其中$\hat{y}_b(x)$是对第$b$颗树的估计。

为了评估估计模型的性能，使用4个指数，参考式（11）~（14）。$F1_{macro}$为宏平均数，用于对所有类别（1, ..., u）加权平均。$F1_{micro}$为微平均数，用于对所有样本平均加权，从而有利于样本预测结果的提升。加权分数在每个标签中找到平均值，然后按每个类的真实实例数加权。Accuracy_{OOB}是训练样本集Z^*的平均准确度，仅使用其引导样本中没有的树[52]。

$$F1_{macro}=\frac{2\times P_{macro}\times R_{macro}}{P_{macro}+R_{macro}}\left(P_{macro}=\frac{1}{u}\sum_{u=1}^{u}\text{Precision}_u,\ R_{macro}=\frac{1}{u}\sum_{u=1}^{u}\text{Recall}_u\right) \tag{11}$$

式中，P_{macro}是宏精确率；R_{macro}是宏召回率。Precision_u表示TP_u数除以TP_u和FP_u的总数（TP_u是类别u中将正类正确预测为正类的数量，FP_u是类别u中将负类错误预测为正类的数量），而Recall_u定义为TP_u数除以TP_u和FN_u的总数（FN_u是类别u中将正类错误预测为负类的数量）。

$$F1_{micro}=\frac{2\times P_{micro}\times R_{micro}}{P_{micro}+R_{micro}}\left(P_{micro}=\frac{\sum_1^u TP_u}{\sum_1^u TP_u+\sum_1^u FP_u},\ R_{micro}=\frac{\sum_1^u TP_u}{\sum_1^u TP_u+\sum_1^u FN_u}\right) \tag{12}$$

式中，P_{micro}是微精确率；R_{micro}是微召回率。

$$\text{Weighted score}=\frac{1}{S}\sum_1^u\frac{2\times\text{Precision}_u\times\text{Recall}_u}{\text{Precision}_u+\text{Recall}_u}\times S_u \tag{13}$$

式中，S_u是类别u中的样本数量；S为样本数量。

$$\text{Accuracy}_{OOB}=\frac{TP_{OOB}+TN_{OOB}}{TP_{OOB}+FP_{OOB}+TN_{OOB}+FN_{OOB}} \tag{14}$$

式中，TP_{OOB}是OOB样本中将正类正确预测为正类的数量，FN_{OOB}是OOB样本中将正类错误预测为负类的数量。FP_{OOB}是OOB样本中将负类错误预测为正类的数量，TN_{OOB}是OOB样本中将负类正确预测为负类的数量。

本研究使用了潜在的时空模式数据。分类问题涉及

识别4个延误程度（“准时”“轻度延误”“中度延误”和“重度延误”）。为了估计模型在实际中的执行效果，本文采用了交叉验证策略。一轮交叉验证涉及将所有记录分为5个互补子集，其中对4个子集执行训练过程，然后在另一个测试集上验证分析。接下来，将验证结果在5个回合中取平均值，以估算模型的性能。总共使用了13 492 326条记录。如上所述，选择每种模式下的飞行概率值（潜在的时空模式）作为特征，即$\{A_1\sim A_4, H_1\sim H_5, W_1\sim W_4, M_1\sim M_4\}$，每个功能的范围是0~1。

树的数量是最重要的变量，其应该足够大以使RF的泛化误差收敛。在图4中，我们发现当B从100增加到150 时，$\text{Accuracy}_{\text{OOB}}$从53.1%变 为53.4%；当$B$大 于150时，$\text{Accuracy}_{\text{OOB}}$稍微增加。在$B$大于150后，RF分类器对$B$的增加几乎不敏感。为了获得更好的参数集，本文使用了网格搜索方法，即通过尝试其他参数的若干种组合来确定此模型的最佳参数值。

各个类别的整体性能和准确性如图5所示。正如先前的研究所指出的，由于延误主要由动态操作因素决定，因此难以在延误发生之前做出准确判断[53,54]。我们的模型仅考虑通过基本飞行信息揭示的潜在模式，在这种情况下，RF的所有性能指标均达到50%以上，各个类别的准确度分别为60.0%、46.0%、44.0%和65.0%，这是一个非常积极的表现。它表示可以仅根据时间和机场信息预先估算不同航班延误的概率。“准时”模式和“重度延误”模式由于具有独特的特征而被更准确地分类。即使该算法在分类“轻度延误”和“中度延误”类别时比较困难，但混淆矩阵对角线附近的深色区域表示错误的估计值具有较小的偏差。

4. 结论

在本文中，我们开发了一个概率分解框架，该框架可将大量飞行记录数据转换为时空高维张量。我们的目的是调查空中交通和航班延误的时空动态模式。我们假设每个飞行观测都是从泛分布中产生的样本。然后，我们使用非负张量因子分解方法进行数据处理。结果表明，清晰的模式可以被挖掘出来。核心张量也有效展示了不同模式的相互作用，解释了延误与时空模式之间的关系。另外，“重度延误”往往发生于每一天的下午，尤其是工作日和夏季的高峰期。从中西部地区机场起飞的航班在一天中的任何时间都很少有延误的可能性。

通过在空间和时间维度上加强对航班的了解，该框架可以为航班延误的建模提供启发。而且，本文已经证明潜在模式对延误有一定作用。随着空间和时间信息的整合，潜在的模式可以给出有关延误程度的估计结果。

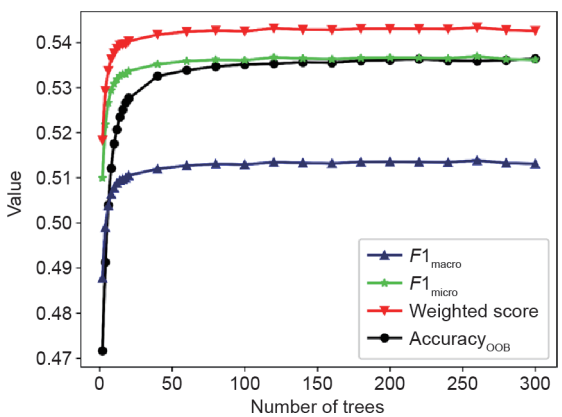

图4. 树的数量对模型性能的影响。

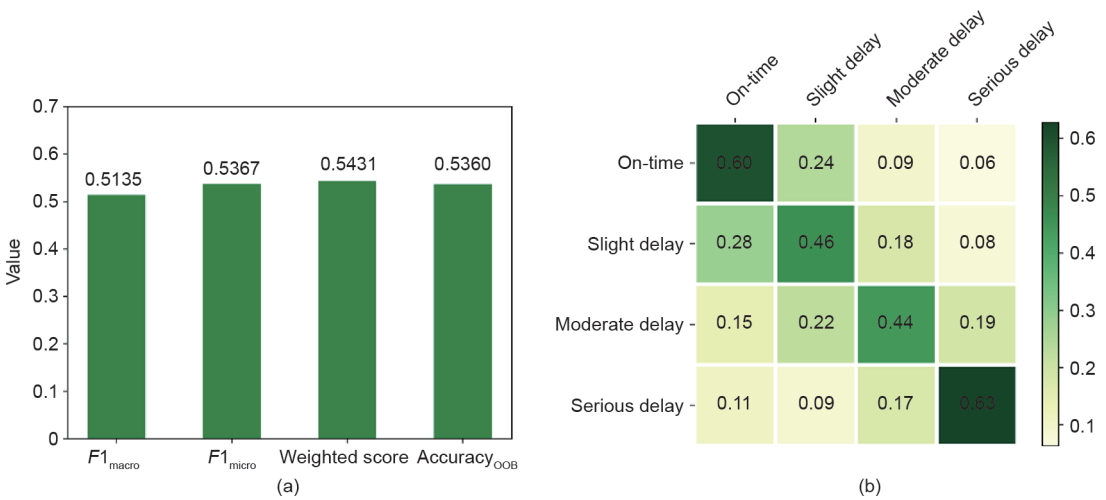

图5. 分类结果。（a）不同指标衡量的总体表现；（b）混淆矩阵。

在高度动态的环境和延误复杂性的背景下，此结果使我们对空中交通和航班延误与时间和空间的相互作用有了新的认识。该框架通过潜在类别模型和概率分解方法对海量航空数据进行深入了解。研究结果可以帮助机场运营商和空中交通管理人员，根据从历史情景中获得的经验，更好地制订空中交通管理策略和完善机场内部管理。未来可以进一步研究涉及更多因素（如天气和航线属性）的相互作用的航班延误问题。

致谢

感谢美国南佛罗里达大学的张瑜博士和美国卡内基梅隆大学的张沈麒对本文撰写工作给予的支持和建议。该研究获得国家重点研发计划（2019YFF0301400）及国家自然科学基金（61671031、61722102和61961146005）资助。

Compliance with ethics guidelines

Mingyuan Zhang, Shenwen Chen, Lijun Sun, Wenbo Du, and Xianbin Cao declare that they have no conflict of interest or financial conflicts to disclose.

References

[1] Folkes VS, Koletsky S, Graham JL. A field study of causal inferences and consumer reaction: the view from the airport. J Consum Res 1987;13 (4):534–9.
[2] Britto R, Dresner M, Voltes A. The impact of flight delays on passenger demand and societal welfare. Transp Res Part E Logist Trans Rev 2012;48(2):460–9.
[3] Ferrer JC, e Oliveira PR, Parasuraman A. The behavioral consequences of repeated flight delays. J Air Transp Manage 2012;20(3):35–8.
[4] Vlachos I, Lin Z. Drivers of airline loyalty: evidence from the business travelers in China. Transp Res Part E Logist Trans Rev 2014;71:1–17.
[5] Cook AJ, Tanner G. European airline delay cost reference values. Report. London: University of Westminster; 2011 Mar.
[6] Pejovic T, Noland RB, Williams V, Toumi R. A tentative analysis of the impacts of an airport closure. J Air Transp Manage 2009;15(5):241–8.
[7] Ryerson MS, Hansen M, Bonn J. Time to burn: flight delay, terminal efficiency, and fuel consumption in the National Airspace System. Transp Res Part A Policy Pract 2014;69:286–98.
[8] Ball M, Barnhart C, Dresner M, Hansen M, Neels K, Odoni A, et al. Total delay impact study: a comprehensive assessment of the costs and impacts of flight delay in the United States. Report. National Center of Excellence for Aviation Operations Research; 2010.
[9] Abdelghany KF, Shah SS, Raina S, Abdelghany AF. A model for projecting flight delays during irregular operation conditions. J Air Transp Manage 2004;10 (6):385–94.
[10] Robinson PJ. The influence of weather on flight operations at the Atlanta Hartsfield International Airport. Weather Forecast 1989;4(4):461–8.
[11] Reynolds-Feighan AJ, Button KJ. An assessment of the capacity and congestion levels at European airports. J Air Transp Manage 1999;5(3):113–34.
[12] Wong JT, Li SL, Gillingwater D. An optimization model for assessing flight technical delay. Transp Plann Technol 2002;25(2):121–53.
[13] Mueller ER, Chatterji G. Analysis of aircraft arrival and departure delay characteristics. In: Proceedings of the AIAA's Aircraft Technology, Integration, and Operations (ATIO) 2002 Technical Forum; 2002 Oct 1–3; Los Angeles, CA, USA; 2002.
[14] Wu CL. Inherent delays and operational reliability of airline schedules. J Air Transp Manage 2005;11(4):273–82.
[15] Schaefer L, Millner D. Flight delay propagation analysis with the detailed policy assessment tool. In: Proceedings of the 2001 IEEE International Conference on Systems, Man and Cybernetics. e-Systems and e-Man for Cybernetics in Cyberspace; 2001 Oct 7–10; Tucson, AZ, USA; 2001.
[16] Han Y, Moutarde F. Analysis of large-scale traffic dynamics in an urban transportation network using non-negative tensor factorization. Int J Intell Transp Syst Res 2016;14:36–49.
[17] Gorripaty S, Liu Y, Hansen M, Pozdnukhov A. Identifying similar days for air traffic management. J Air Transp Manage 2017;65:144–55.
[18] Hoffman B, Krozel J, Penny S, Roy A, Roth K. A cluster analysis to classify days in the National Airspace System. In: Proceedings of the AIAA Guidance, Navigation, and Control Conference and Exhibit; 2003 Aug 11–14; Austin, TX, USA; 2003.
[19] Liu Y, Seelhorst M, Pozdnukhov A, Hansen M, Ball MO. Assessing terminal weather forecast similarity for strategic air traffic management. In: Proceedings of the 6th International Conference on Research in Air Transportation; 2014 May 26–30; Istanbul, Turkey; 2014.
[20] Mukherjee A, Grabbe SR, Sridhar B. Classification of days using weather impacted traffic in the National Airspace System. In: Proceedings of the 2013 Aviation Technology, Integration, and Operations Conference; 2013 Aug 12–14; Los Angeles, CA, USA; 2013.
[21] Grabbe SR, Sridhar B, Mukherjee A. Clustering days with similar airport weather conditions. In: Proceedings of the 14th AIAA Aviation Technology, Integration, and Operations Conference; 2014 Jun 16–20; Atlanta, GA, USA; 2014.
[22] Bloem M, Bambos N. Ground delay program analytics with behavioral cloning and inverse reinforcement learning. In: Proceedings of the 14th AIAA Aviation Technology, Integration, and Operations Conference; 2014 Jun 16–20; Atlanta, GA, USA; 2014.
[23] Sternberg A, Carvalho D, Murta L, Soares J, Ogasawara E. An analysis of Brazilian flight delays based on frequent patterns. Transp Res Part E Logist Trans Rev 2016;95:282–98.
[24] Zhou G, Cichocki A, Zhao Q, Xie S. Efficient nonnegative tucker decompositions: algorithms and uniqueness. IEEE Trans Image Process 2015;24(12):4990–5003.
[25] Abdel-Aty M, Lee C, Bai Y, Li X, Michalak M. Detecting periodic patterns of arrival delay. J Air Transp Manage 2007;13(6):355–61.
[26] Zhou X, List GF. An information-theoretic sensor location model for traffic origin-destination demand estimation applications. Transp Sci 2010;44 (2):254–73.
[27] Huang J, Levinson D, Wang J, Zhou J, Wang ZJ. Tracking job and housing dynamics with smartcard data. Proc Natl Acad Sci USA 2018;115(50):12710–5.
[28] Du WB, Zhang MY, Zhang Y, Cao XB, Zhang J. Delay causality network in air transport systems. Transp Res Part E Logist Trans Rev 2018;118:466–76.
[29] Mislevy RJ. Estimating latent distributions. Psychometrika 1984;49 (3):359–81.
[30] Woodbury MA, Manton KG. Grade of membership analysis of depression-related psychiatric disorders. Sociol Methods Res 1989;18(1):126–63.
[31] Sun L, Axhausen KW, Lee DH, Huang X. Understanding metropolitan patterns of daily encounters. Proc Natl Acad Sci USA 2013;110(34):13774–9.
[32] Zhang F, Wilkie D, Zheng Y, Xie X. Sensing the pulse of urban refueling behavior. In: Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing; 2013 Sep 8–12; Zurich, Switzerland; 2013. p. 13–22.
[33] Yuan J, Zheng Y, Xie X. Discovering regions of different functions in a city using human mobility and POIs. In: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2012 Aug 12–16; Beijing, China; 2012. p. 186–94.
[34] Ma X, Wu YJ, Wang Y, Chen F, Liu J. Mining smart card data for transit riders' travel patterns. Transp Res Part C Emerg Technol 2013;36:1–12.
[35] Dauwels J, Aslam A, Asif MT, Zhao X, Vie NM, Cichocki A, et al. Predicting traffic speed in urban transportation subnetworks for multiple horizons. In: Proceedings of the 13th International Conference on Control Automation Robotics & Vision; 2014 Dec 10–12; Singapore; 2014. p. 547–52.
[36] Ran B, Tan H, Wu Y, Jin PJ. Tensor based missing traffic data completion with spatial–temporal correlation. Physica A Stat Mech Its Appl 2016;446:54–63.
[37] Asif MT, Mitrovic N, Dauwels J, Jaillet P. Matrix and tensor based methods for missing data estimation in large traffic networks. IEEE Trans Intell Transp Syst 2016;17(7):1816–25.
[38] Liu C, Chen X. Vessel track recovery with incomplete AIS data using tensor CANDECOM/PARAFAC decomposition. J Navig 2014;67(1):83–99.
[39] Gaussier E, Goutte C. Relation between PLSA and NMF and implications. In: Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval; 2005 Aug 15–19; Salvador, Brazil; 2005. p. 601–2.
[40] Tucker LR. Implications of factor analysis of three-way matrices for measurement of change. In: Harris CW, editor. Problems in measuring change. Madison: University of Wisconsin Press; 1963. p. 122–37.
[41] De Lathauwer L, De Moor B, Vandewalle J. A multilinear singular value decomposition. SIAM J Matrix Anal Appl 2000;21(4):1253–78.

[42] Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. Nature 1999;401(6755):788–91.
[43] Hagenaars JA, McCutcheon AL. Applied latent class analysis. Cambridge: Cambridge University Press; 2002.
[44] McCutcheon AL. Latent class analysis. Newbary Park: Sage Publications, Inc.; 1987.
[45] Peng W, Li T. On the equivalence between nonnegative tensor factorization and tensorial probabilistic latent semantic analysis. Appl Intell 2011;35 (2):285–95.
[46] Ding C, Li T, Peng W. On the equivalence between non-negative matrix factorization and probabilistic latent semantic indexing. Comput Stat Data Anal 2008;52(8):3913–27.
[47] McLachlan GJ, Krishnan T. The EM algorithm and extensions. 2nd ed. Hoboken: John Wiley & Sons; 2008.
[48] Sun L, Axhausen KW. Understanding urban mobility patterns with a probabilistic tensor factorization framework. Transp Res Part B Methodol 2016;91:511–24.
[49] Hao L, Hansen M, Zhang Y, Post J. New York, New York: two ways of estimating the delay impact of New York airports. Transp Res Part E Logist Trans Rev 2014;70:245–60.
[50] Breiman L. Random forests. Mach Learn 2001;45:5–32.
[51] Cutler DR, Edwards TC Jr, Beard KH, Cutler A, Hess KT, Gibson J, et al. Random forests for classification in ecology. Ecology 2007;88(11): 2783–92.
[52] Gislason PO, Benediktsson JA, Sveinsson JR. Random forests for land cover classification. Pattern Recognit Lett 2006;27(4):294–300.
[53] Rebollo JJ, Balakrishnan H. Characterization and prediction of air traffic delays. Transp Res Part C Emerg Technol 2014;44:231–41.
[54] Sternberg A, Soares J, Carvalho D, Ogasawara E. A review on flight delay prediction. 2017. arXiv:1703.06118.