

基于策略学习的机票动态定价算法

卢敏^{*①} 张耀元^① 卢春^②

^①(中国民航大学计算机科学与技术学院 天津 300300)

^②(中国南方航空股份有限公司信息中心 广州 510000)

摘要: 机票动态定价旨在构建机票售价策略以最大化航班座位收益。现有机票定价算法都建立在提前预测各票价等级的需求量基础之上, 会因票价等级需求量的预测偏差而降低模型性能。为此, 提出基于策略学习的机票动态定价算法, 其核心是不再预测各票价等级的需求量, 而是将机票动态定价问题建模为离线强化学习问题。通过设计定价策略评估和策略更新的方式, 从历史购票数据上学习具有最大期望收益的机票动态定价策略。同时设计了与现行定价策略和需求量预测方法的对比方法及评价指标。在两趟航班的多组定价结果表明: 相比于现行机票销售策略, 策略学习算法在座位收益上的提升率分别为30.94%和39.96%, 且比基于需求量预测方法提升了6.04%和3.36%。

关键词: 民航收益管理; 机票动态定价; 强化学习; 策略学习

中图分类号: TP311

文献标识码: A

文章编号: 1009-5896(2021)04-1022-07

DOI: 10.11999/JEIT200778

Approach for Dynamic Flight Pricing Based on Strategy Learning

LU Min^① ZHANG Yaoyuan^① LU Chun^②

^①(College of Computer Science and Technology, Civil Aviation University of China, Tianjin 300300, China)

^②(Information Center of China Southern Air Holding Company Limited, Guangzhou 510000, China)

Abstract: The core of the dynamic flight pricing is to yield a pricing strategy with maximum seat revenue. The state-of-the-art flight pricing approaches are built on forecasting the fare demand. They suffer low profit due to the inaccurate prediction. To tackle the above issue, an approach for dynamic flight pricing based on strategy learning is proposed. That approach resorts to reinforcement learning to output pricing strategy with the highest expected return. That strategy is learned by iteratively policy evaluation and policy improvement. The rate of profit improvement on the two flights is empirically 30.94% and 39.96% over the existing pricing strategy, while that rate is 6.04% and 3.36% over the demand forecasting algorithm.

Key words: Revenue management; Dynamic flight pricing; Reinforcement learning; Strategy learning

1 引言

航空公司采用收益管理技术提升了2%至8%的票务年收入^[1]。机票动态定价是收益管理的核心, 旨在构建航班机票售价策略, 将不同票价等级的机票适时地销售给不同的消费者, 以实现收益最大化。

Gallego等人^[2]使用强度控制理论求解了不确定需求下预期收入的上限。Otero等人^[3]采用阶段类型分布和更新过程建模客户购买机票的时间及购买

概率。Delahaye等人^[4]通过细分商务和休闲旅客的不同价格敏感度进行动态定价。高金敏等人^[5]研究了时变需求下动态定价模型。Selcuk等人^[6]针对4种不同的价格需求关系提出了精确的动态规划方法。除了上述面向垄断市场环境的定价研究^[1-6], 还有从竞争角度出发的定价研究^[7,8]。Lin等人^[7]构建了基于博弈论的实时价格预测模型。施飞等人^[8]通过航空公司与旅客的动态博弈解决各等级票价的售票时间。

现有机票动态定价方法都建立在提前预测各票价等级的需求量基础之上。然而, 实际应用中各等级票价的需求量受节假日、季节、售票时机等诸多因素影响。基于历史销售数据进行函数拟合的需求量预测方法存在预测偏差, 降低了建立在需求量预测方法之上的模型性能。为此, 本文提出基于策略学习的机票动态定价算法。其核心是不再直接预测

收稿日期: 2020-09-20; 改回日期: 2021-02-04; 网络出版: 2021-03-02

*通信作者: 卢敏 ml@cauc.edu.cn

基金项目: 国家自然科学基金(61502499), 民航航空公司人工智能重点实验室项目

Foundation Items: The National Natural Science Foundation of China (61502499), The Project from Key Laboratory of Artificial Intelligence for Airlines, CAAC

各票价等级的需求量，而是从历史购票数据上学习具有最大期望收益的机票定价策略。在两趟航班两年历史销售数据上的多组模拟定价结果表明：相比于现行定价策略，策略学习算法的票务收益平均提升率分别为30.94%和39.96%，且比需求量预测方法提升了6.04%和3.36%。

目前存在将强化学习应用于收益管理领域的研究，如电动汽车^[9]、广告竞价^[10]、可再生能源管理^[11]、机票超售及座位控制^[12-15]。Gosavii等人^[12]提出了面向超售和多等级票价的半马尔可夫决策模型。Shihab等人^[13]采用深度强化学习优化航班座位控制策略。Qiu等人^[14]建模了旅客行为驱动的航班定价。Lawhead等人^[15]提出一种航班座位收益有界的动作评论算法。不同于上述工作，本文聚焦于采用离线强化学习求解机票动态定价策略。

2 航班机票动态定价算法

在机票销售期内以某票价等级出售机票，售票系统会得到立即收益，即座位销售数×票价等级×机票全价，同时减少航班剩余座位数。为了最大化后期座位收益，航班剩余座位数又影响下一时刻的机票定价。为此，机票动态定价受航班剩余座位数、定价策略和收益等相互影响，可建模为强化学习问题。

航班机票动态定价的强化学习要素分别为：(1)状态集 \mathbf{S} 表示航班剩余座位数。设 n 为某航班座位容量，航班在时刻 t 的状态(剩余座位数) $s_t \in \mathbf{S} = \{n, n-1, \dots, 0\}$ 。(2)动作集 \mathbf{A} 表示所有的票价等级，本质为经济舱全票价的折扣率。航班在时刻 t 采用的动作(票价等级)为 $\alpha_t \in \mathbf{A}$ 。(3)策略函数 $\pi(\alpha|s)$ 表示在状态 s 下执行动作 α 的概率。策略函数是算法的学习目标，使得算法根据它产生动作序列。(4)收益函数 $r(s, \alpha)$ 表示在状态 s 下执行动作 α 的立即收益，取值为 $p \times \alpha \times N$ ，其中 p, N 分别为经济舱全价和销售的座位数。

引入动作值函数 $q_\pi(s, \alpha)$ 以评估策略函数 $\pi(\alpha|s)$ 的好坏程度，见式(1)。式(1)的物理含义是给定状态 s 沿着策略 π 执行序贯性决策后预期总折扣奖励。

$$q_\pi(s, \alpha) = E_\pi \left[\sum_{t=0}^{+\infty} \gamma^t r(s_t, \alpha_t) \mid s_0 = s, \alpha_0 = \alpha \right] \quad (1)$$

其中， $E_\pi[\cdot]$ 表示在策略 π 下的期望收益。 $\gamma \in [0, 1]$ 表示收益折扣因子，决定了未来收益的现值，即 $t+1$ 时刻的收益只有当前收益的 γ^t 倍。学习目标是构建具有最大值 $q_\pi(s, \alpha)$ 的最优策略 π^* ，使得对于任何状态 s 都有 $q(s, \pi^*(s)) \geq q_\pi(s, \alpha)$ 。此时最优策略 π^* 的Bellman方程展开为

$$q_{\pi^*}(s, \alpha) = \sum_{s_{t+1}} \Pr(s_{t+1} \mid s_t = s, \alpha_t = \alpha) \times \left[r(s_t, \alpha_t) + \gamma \max_{\alpha'} q(s_{t+1}, \alpha') \right] \quad (2)$$

其中， $\Pr(s_{t+1} \mid s_t = s, \alpha_t = a)$ 表示在状态 s_t 下执行动作 a_t 后跳转到状态 s_{t+1} 的概率。由于机票销售过程无法建模上述转移概率而仅可观察剩余座位数 s_{t+1} ，采用时序差分法^[16]更新动作值函数。

$$q_{\pi^*}(s_t, a_t) = q_{\pi^*}(s_t, a_t) + \eta \left(r(s_t, \alpha_t) + \gamma \max_{a_{t+1}} q_{\pi^*}(s_{t+1}, a_{t+1}) - q_{\pi^*}(s_t, a_t) \right) \quad (3)$$

其中， η 表示学习速率。当动作值函数通过更新达到收敛后，每个状态 s 在最优策略 π^* 的动作 $\pi^*(s)$ 为

$$\pi^*(s) = \arg \max_{\alpha} q(s, \alpha) \quad (4)$$

由于航班动态定价的状态空间和动作空间都离散且有限，提出基于策略学习的机票动态定价算法，见表1。算法首先初始化一个2维矩阵 \mathbf{Q} -table以表示动作值函数 $q(s, \alpha)$ ，其中矩阵的行数为状态空间数 $|\mathbf{S}|$ ，列数是动作空间数 $|\mathbf{A}|$ 。其次，在航班的第1天到 $T-1$ 天的历史销售序列上依次据式(3)和式(4)进行策略 π 的评估和改进。重复上述循环操作，直至算法达到预先设定的最大迭代次数。

3 实验

3.1 实验设置

3.1.1 数据集

实验数据集是航班历史机票销售序列，是由

表1 机票动态定价策略学习算法

输入	学习速率 η ，折扣因子 γ ，最大迭代次数episodes，航班总座位数 N ，航班第1天到 $T-1$ 天的历史销售序列 $\{s_0^{(n)}, a_0^{(n)}, r_0^{(n)}, \dots, s_v^{(n)}, a_v^{(n)}, r_v^{(n)}\}_{n=1}^{T-1}$
初始化	对于任何状态 s 和 $\alpha, q(s, \alpha) = 0, k = 0, n = 1$
Repeat:	
Repeat	(对于第1天到 $T-1$ 天的每趟离港航班):
Repeat	(对于此趟航班历史销售序列的每一步 $(s_t^{(n)}, a_t^{(n)}, r_t^{(n)}, s_{t+1}^{(n)})$):
策略评估:	据式(3)更新动作值函数 $q(s_t^{(n)}, a_t^{(n)})$
策略更新:	按式(4)调整策略 $\pi(s_t^{(n)}) = \arg \max_{\alpha} q(s_t^{(n)}, \alpha)$
Until	航班没有剩余座位或售票时间截止
$n \leftarrow n + 1$	
Until	$n > T - 1$
$k \leftarrow k + 1$	
Until	$k > \text{episodes}$
输出	第 T 天的机票动态定价策略 $\pi(s) = \arg \max_{\alpha} q(s, \alpha)$

2010年1月1日到2011年12月31日的民航旅客订票日志PNR (Passenger Name Record)生成的。一条PNR是一次订票信息,主要字段包括旅客身份证号码(信息加密)、航空公司、航班号、出发机场和到达机场(机场三字码表示)、订单编号、票价等级等,示例见表2。

在民航旅客订票日志上第 n 天离港航班的机票销售序列生成过程为:首先按照组合字段(航班号,离港日期)筛选出此航班在第 n 天的订票记录,其次对筛选的记录按订单号升序排列,然后将具有相同订单号的订票记录合并以生成一个四元组 $(s_t^{(n)}, a_t^{(n)}, r_t^{(n)}, s_{t+1}^{(n)})$ 。其中 $s_t^{(n)}$ 表示第 n 天离港航班在时刻 t 销售前的航班剩余座位数; $a_t^{(n)}$ 为当前订单的票价等级; $r_t^{(n)}$ 表示执行动作 $a_t^{(n)}$ 的收益,取值为当前订单机票销售数 \times 票价等级 \times 机票全价; $s_{t+1}^{(n)}$ 表示时刻 t 销售后的航班剩余座位数。重复上述过程,直至生成航班机票销售序列。

折中于算法在民航旅客订票日志上所有航班的运行时间,以及评价航班每一天动态定价策略,挑选在航线热度、经济舱总数、实际销售等方面差异较大的两个航班CA1501和JR1505进行实验,其统计信息见表3。民航旅客订票日志中票价等级的精度为0.0001,且各票价等级销售的机票数比较稀疏。意味着很多二元组(剩余座位数,票价等级)未在历史数据中出现,使得这些元组对应的动作值函数无法更新,导致后期动态定价存在较大的偏差。比如航班CA1501历史销售轨迹中产生了5737种票价等级,而各票价等级产生平均机票数为22.68,这表明动作值函数矩阵 Q -table中某一列平均至少有 $255-22.68$ 个元素未被更新,累计有 $5737 \times (255-22.68)$ 个元素未更新,其中255是航班CA1501的总座位数。精确度为千分位也会产生类似的结论。为此,实验中将票价等级精度降至百分

位,并在实验中分析万分位精度对航班座位收益的影响。

3.1.2 评价指标

为评价航班动态定价策略好坏程度,定义航班收益平均提升率ALR@T(Average Lift Rate at T days):

$$ALR@T = \frac{1}{T-1} \sum_{n=2}^T \frac{G_{\pi}^{(n)} - G_g^{(n)}}{G_g^{(n)}} \times 100\% \quad (5)$$

其中,ALR@T计算从第2天到第 T 天航班收益提升率的均值。从第2天开始实验评价的原因是:第1天没有历史销售序列导致算法无法学习第1天的航班定价策略; $(G_{\pi}^{(n)} - G_g^{(n)})/G_g^{(n)}$ 是第 n 天航班的收益提升率; $G_{\pi}^{(n)}$ 表示根据第 n 天动态定价策略 π 产生的收益; $G_g^{(n)}$ 则是基准算法在第 n 天航班上定价策略的收益。

3.1.3 基准算法

(1) 航班现行定价策略

历史销售序列本质反映航班现行定价策略。为了能与现行策略比较,动态定价策略 π 采取类似方法售票:在第 n 天离港航班的时刻 t ,若现行策略以票价等级 $a_t^{(n)}$ 销售 $N_t^{(n)}$ 张机票,则动态定价策略 π 在时刻 t 状态 $s_t^{(n)}$ 会以票价等级 $\pi(s_t^{(n)})$ 销售 $N_t^{(n)}$ 张机票。此时现行策略的收益 $G_g^{(n)}$ 和定价策略 π 的收益 $G_{\pi}^{(n)}$ 分别为

$$G_g^{(n)} = \sum_t a_t^{(n)} N_t^{(n)}, \quad G_{\pi}^{(n)} = \sum_t \pi(s_t^{(n)}) N_t^{(n)} \quad (6)$$

(2) 票价等级需求量预测方法

票价等级需求量预测方法利用前 $n-1$ 天机票销售记录,预测航班第 n 天各票价等级的需求量。在软件1stOpt上利用麦夸特法分别拟合航班CA1501和JR1505的票价等级需求函数,平均拟合优度为94.28%和92.37%。设航班第 n 天各票价等级

表2 旅客订票记录示例

身份证号	航空公司	航班号	出发机场	到达机场	出发日期	订单编号	票价等级
52893787	CA	1501	PEK	SHA	20100308	2273651247	0.5213
55503718	CA	1501	PEK	SHA	20100308	2745812364	0.8212

表3 实验数据集的统计信息

航班	售票记录总数	销售序列数	状态、动作等四元组数	原始票价等级(精确到万分位)		预处理后的票价等级(精确到千分位)		预处理后的票价等级(精确到百分位)	
				票价等级数	各等级平均票数	票价等级数	各等级平均票数	票价等级数	各等级平均票数
CA1501	130118	718	102809	5737	22.68	1087	119.70	150	867.45
JR1505	22691	611	17102	2359	9.62	745	30.46	90	254.96

$\alpha_1, \dots, \alpha_{|A|}$ 的需求量为 $n_1, \dots, n_{|A|}$ ，需求量预测方法的收益 $G_g^{(n)} = \sum_{i=1}^{|A|} a_i n_i$ 。

由于需求量预测方法无法产生机票销售序列，而策略 π 却需根据剩余座位数选择票价等级，两者无法直接对比。为了能够对比，策略 π 基于如下假设售票：(a)将航班机票售卖期划分为 m 个离散周期 $l = 1, 2, \dots, m$ ；(b)各周期内旅客购买票价等级 α_i 的数量服从均值为 n_i/m 的泊松分布；(c)每次仅售1张机票。使用 h_{li} 表示第 l 周期中票价等级 α_i 的销售量，第 l 周期总销售量为 $h_l = \sum_{i=1}^{|A|} h_{li}$ 。两定价策略在第1张机票销售收益差 Δ_1 ：

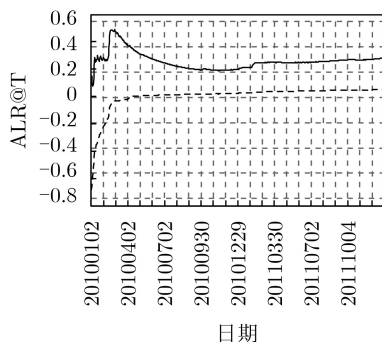
$$\Delta_1 = \arg \max_a q(s_0^{(n)}, a) - \left[\frac{h_{11}}{h_1} a_1 + \frac{h_{12}}{h_1} a_2 + \dots + \frac{h_{1|A|}}{h_1} a_{|A|} \right] \quad (7)$$

其中， $\arg \max_a q(s_0^{(n)}, a)$ 表示策略 π 在状态 $s_0^{(n)}$ 下采取动作 $a = \pi(s_0^{(n)})$ 的收益。 $\sum_{i=1}^{|A|} h_{1i} \alpha_i / h_1$ 是需求量预测方法的第一张机票平均收益，其计算原理是：因无法建模机票销售顺序，假设所有销售序列都是等可能的，则第1周期内第1张机票以票价等级 $\alpha_1, \dots, \alpha_{|A|}$ 出售概率是 $h_{11}/h_1, \dots, h_{1|A|}/h_1$ 。依式(7)类推，第1周期内总收益差为

$$\sum_{i=1}^{h_1} \Delta_i = \sum_{s=|S|-h_1+1}^{|S|} \arg \max_a q(s, a) - [h_{11} a_1 + h_{12} a_2 + \dots + h_{1|A|} a_{|A|}] \quad (8)$$

其中， $|S|$ 为航班总座位数。依式(8)计算 m 个周期的收益差，需求量预测方法和策略 π 在第 n 天航班收益差为

$$G_\pi^{(n)} - G_g^{(n)} = \sum_{s=|S|-(n_1+n_2+\dots+n_{|A|})+1}^{|S|} \arg \max_a q(s, a) - \sum_{i=1}^{|A|} a_i n_i \quad (9)$$



(a) 航班CA1501实验结果

3.1.4 参数设置和运行环境

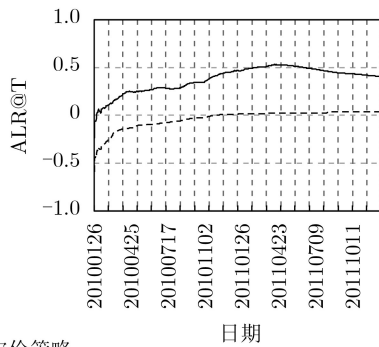
航班CA1501和JR1505经济舱总座位数 N 分别为255和52。算法需预先设置的参数分别为：学习速率 $\eta = 0.2, 0.4, 0.6, 0.8$ 、折扣因子 $\gamma = 1$ 、算法总迭代次数 $\text{episodes} = 5000$ 。算法硬件环境为Intel(R) Xeon(R) Silver 4110、主频3.4 GHz、内存126 GB。实验训练共耗约142 h。

3.2 实验对比及分析

3.2.1 航班 CA1501实验结果

在航班CA1501进行了717组实验，其结果见图1(a)。相比于现行定价策略，动态定价策略 π 的收益平均提升率 ALR@T 为30.94%；相比票价等级需求量预测方法，平均提升率为6.04%。

图1(a)中动态定价策略 π 相比于现行定价策略的提升率曲线——策略 π vs现行定价策略——可分为4段。在2010年1月2日到2010年2月10日，1月1日至1月3日票务收益较差，使得学习后的策略收益也低。1月4日收益较1月3日增长了115%，算法能快速学习此定价策略以用于后续定价。故此阶段内曲线出现骤增，平均提升率 ALR@T 稳定至约35%。在2010年2月11日到2010年2月19日曲线出现第2次骤增，其原因是：2010年2月11日为农历腊月二十八，出现“倒春运”，即机票销售数和价格都明显下跌，期间平均收益仅为整体的32.29%。然而，定价策略 π 是从2月11日之前销售数据上学习，导致策略 π 的机票价格要高于“倒春运”期间价格，出现较大收益差值。第3段为2010年2月11日到2011年1月30日，曲线逐渐降低并趋于平稳，收敛至约21%。其原因是：在大量历史数据上的学习，动态定价策略与现行定价策略的收益差值逐渐减小。最后一次曲线提升发生在2011年2月1日，其原因是：2011年2月1日为农历腊月二十九，也出现“倒春运”，因在大量历史数据上的学习避免了类似第1次突变，趋于平稳。



(b) 航班JR1505实验结果

图1 两个航班上定价策略的性能对比

图1(a)也列出了定价策略 π 相比于需求量预测方法的提升率曲线——策略 π vs需求量预测方法。实验初期因训练数据较少,动态定价策略产生的收益低于需求量预测方法。随着训练数据的增加,策略学习方法与需求量预测方法的收益差距逐渐减小,曲线平稳提升并趋于稳定,最终收益平均提升率为6.04%。

3.2.2 航班JR1505实验结果

相比于航班CA1501,航班JR1505收益更易受定价影响。航班JR1505上的610组实验结果见图1(b)。动态定价策略 π 相比于现行定价策略和需求量预测方法的收益平均提升率ALR@T分别为39.96%和3.36%。

从图1(b)可知,动态定价策略 π 相比于现行定价策略的提升率曲线——策略 π vs现行定价策略,主要可分为3段。2010年1月26日到2010年11月10日为第1个增长期。定价策略 π 通过不断学习以往经验,平均提升率ALR@T逐渐提升,在4月20日趋于平稳,收敛至25%。第2个增长周期发生在2010年11月2日,其原因是:航班JR1505在此后一个月机票销售收益较前一月平均降低了43.80%,使得收益差值增加,曲线逐渐上升。第3个拐点发生在2011年4月20日,其原因是:航班由淡季转为旺季,此后一月较上月收益提升41.22%。实际收益的提升使得差值逐渐减小,曲线回落并重新收敛至39%。此外,相比于需求量预测方法,策略学习方法随着不断学习,平均提升率ALR@T平稳提升并趋于稳定,最终收敛至约3%。

3.3 票价等级精度对收益的影响

表4表明不同的票价等级精度产生不同规模的票价等级,为此分析票价等级精度对算法性能的影响。表4列出2011年6月22日航班CA1501在票价等级精度为0.0001和0.0100的实验结果。万分位精度产生了4590个动作,而百分位精度缩减至120个动

作。百分位精度较万分位在平均提升率ALR@T上提升3.17%。

分析算法在精度为0.0100时性能提升的原因。图2(a)列出票价等级精度分别为0.0100和0.0001时定价策略 π 在各状态上的票价等级差值。统计表明,不同精度下的模型在所有状态上采取的票价等级总体相差不大,票价等级差的均值为0.05,方差为0.006。仍有部分点波动较大,造成收益差距。图2(b)以波动点——剩余座位数196为例分析。当精度为0.0001时,采取的票价等级为0.8453,落在区间(0.84,0.85]频数为8。当精度为0.0100时,采取的票价等级为1.04,落在区间(1.03,1.04]频数为65。为此,精度为0.0100更易捕捉具有明显集中趋势的“众数”票价等级,以代表此状态下整体票价水平,而非选择出现频次最高的单一票价等级。

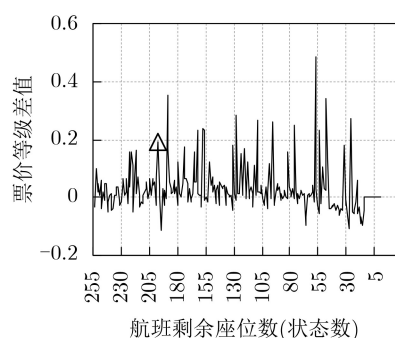
3.4 学习速率的影响分析

学习速率 η 影响着算法的收敛速度和定价策略的收益,以2011年12月31日航班CA1501为例分析学习速率的影响。图3(a)指出学习速率为0.2, 0.4, 0.6, 0.8时,算法收敛的迭代次数依次为4860, 2167, 1274, 788。因此,经验上学习速率在0.2以上就能保证算法收敛。图3(b)指出算法在5000次迭代时,增大学习率不会对收益平均提升率ALR@T产生影响,其中计算ALR@T的基准算法为航班现行定价策略。

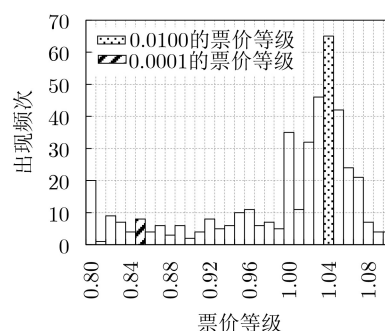
上述经验分析指出学习速率 $\eta = 0.2, 0.4, 0.6, 0.8$ 都是可取范围,在实验过程中学习速率取值为0.6,其原因:(1)折中于算法整体运行时间。由于航班CA1501和JR1505分别需要进行717和610组实

表4 票价等级精度影响分析

票价等级精度	训练集中票价等级总数	定价策略中出现票价等级总数	收益平均提升率ALR@T(%)
0.0001	4590	128	13.21
0.0100	120	16	16.38



(a) 百分位与万分位精度在各状态下的票价等级差值



(b) 剩余票数196时的历史票价等级统计直方图

图2 2011年6月22日航班CA1501在精度为0.0100和0.0001下的实验对比

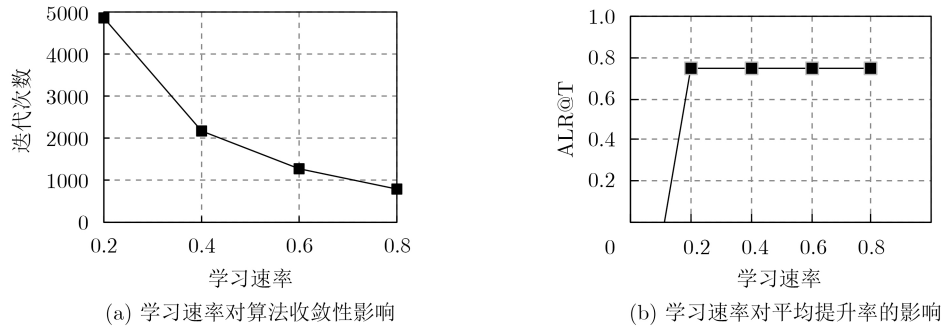


图3 学习速率对算法性能的影响

验，而每组实验的运行时间正比于算法迭代次数，当学习速率 $\eta = 0.6$ 时，算法能够更快收敛。(2)突出同一趟航班上离起飞日期越近的座位票价影响要略大于较早售出座位的票价。当将式(3)重写为 $q_{\pi^*}(s_t, a_t) = (1 - \eta)q_{\pi^*}(s_t, a_t) + \eta(r(s_t, \alpha_t) + \gamma \max_{a_{t+1}} q_{\pi^*}(s_{t+1}, a_{t+1}))$ 时，可知学习速率 η 越大，保留之前训练的效果就越少。如学习速率为1，则训练过程中只考虑新销售的座位的票价，不保留之前训练的效果，不符合实际。

4 结论

针对航班票价定价过程中航班剩余座位数与机票价格相互影响，提出了基于策略学习的机票动态定价算法。算法通过定价策略评估和策略更新，输出具有最大期望收益的机票动态定价策略。在海量民航旅客订票日志上两个航班的多组模拟定价结果表明：相比于现行机票销售策略和需求量预测方法，算法在收益平均提升率上有显著的提升。后续将对航班状态的时序特征展开研究。

参考文献

- [1] SMITH B C, LEIMKUHNER J F, and DARROW R M. Yield management at American airlines[J]. *Interfaces*, 1992, 22(1): 8-31. doi: [10.1287/inte.22.1.8](https://doi.org/10.1287/inte.22.1.8).
- [2] GALLEGO G and VAN RYZIN G. Optimal dynamic pricing of inventories with stochastic demand over finite horizons[J]. *Management Science*, 1994, 40(8): 999-1020. doi: [10.1287/mnsc.40.8.999](https://doi.org/10.1287/mnsc.40.8.999).
- [3] OTERO D F and AKHAVAN-TABATABAEI R. A stochastic dynamic pricing model for the multiclass problems in the airline industry[J]. *European Journal of Operational Research*, 2015, 242(1): 188-200. doi: [10.1016/j.ejor.2014.09.038](https://doi.org/10.1016/j.ejor.2014.09.038).
- [4] DELAHAYE T, ACUNA-AGOST R, BONDOUX N, et al. Data-driven models for itinerary preferences of air travelers and application for dynamic pricing optimization[J]. *Journal of Revenue and Pricing Management*, 2017, 16(6): 621-639. doi: [10.1057/s41272-017-0095-z](https://doi.org/10.1057/s41272-017-0095-z).
- [5] 高金敏, 乐美龙, 曲林迟, 等. 基于时变需求的机票动态定价研究[J]. *南京航空航天大学学报*, 2018, 50(4): 570-576. doi: [10.16356/j.1005-2615.2018.04.020](https://doi.org/10.16356/j.1005-2615.2018.04.020).
GAO Jinmin, LE Meilong, QU Linchi, et al. Dynamic pricing of air tickets based on time-varying demand[J]. *Journal of Nanjing University of Aeronautics & Astronautics*, 2018, 50(4): 570-576. doi: [10.16356/j.1005-2615.2018.04.020](https://doi.org/10.16356/j.1005-2615.2018.04.020).
- [6] SELÇUK A M and AVŞAR Z M. Dynamic pricing in airline revenue management[J]. *Journal of Mathematical Analysis and Applications*, 2019, 478(2): 1191-1217. doi: [10.1016/j.jmaa.2019.06.012](https://doi.org/10.1016/j.jmaa.2019.06.012).
- [7] LIN K Y and SIBDARI S Y. Dynamic price competition with discrete customer choices[J]. *European Journal of Operational Research*, 2009, 197(3): 969-980. doi: [10.1016/j.ejor.2007.12.040](https://doi.org/10.1016/j.ejor.2007.12.040).
- [8] 施飞, 陈森发. 随时间变化的机票折扣定价研究[J]. *交通运输系统工程与信息*, 2010, 10(1): 112-116. doi: [10.3969/j.issn.1009-6744.2010.01.017](https://doi.org/10.3969/j.issn.1009-6744.2010.01.017).
SHI Fei and CHEN Senfa. Air ticket discount pricing based on time varying[J]. *Journal of Transportation Systems Engineering and Information Technology*, 2010, 10(1): 112-116. doi: [10.3969/j.issn.1009-6744.2010.01.017](https://doi.org/10.3969/j.issn.1009-6744.2010.01.017).
- [9] LEE J, LEE E and KIM J. Electric vehicle charging and discharging algorithm based on reinforcement learning with data-driven approach in dynamic pricing scheme[J]. *Energies*, 2020, 13(8): 1950. doi: [10.3390/en13081950](https://doi.org/10.3390/en13081950).
- [10] CHENG Yin, ZOU Luobao, ZHUANG Zhiwei, et al. An extensible approach for real-time bidding with model-free reinforcement learning[J]. *Neurocomputing*, 2019, 360: 97-106. doi: [10.1016/j.neucom.2019.06.009](https://doi.org/10.1016/j.neucom.2019.06.009).
- [11] 陈前斌, 谭颀, 魏延南, 等. 异构云无线接入网架构下面向混合能源供应的动态资源分配及能源管理算法[J]. *电子与信息学报*, 2020, 42(6): 1428-1435. doi: [10.11999/JEIT190499](https://doi.org/10.11999/JEIT190499).
CHEN Qianbin, TAN Qi, WEI Yannan, et al. Dynamic resource allocation and energy management algorithm for hybrid energy supply in heterogeneous cloud radio access

- networks[J]. *Journal of Electronics & Information Technology*, 2020, 42(6): 1428–1435. doi: [10.11999/JEIT190499](https://doi.org/10.11999/JEIT190499).
- [12] GOSAVII A, BANDLA N, and DAS T K. A reinforcement learning approach to a single leg airline revenue management problem with multiple fare classes and overbooking[J]. *IIE Transactions*, 2002, 34(9): 729–742. doi: [10.1080/07408170208928908](https://doi.org/10.1080/07408170208928908).
- [13] SHIHAB S A M, LOGEMANN C, THOMAS D G, *et al.* Autonomous airline revenue management: A deep reinforcement learning approach to seat inventory control and overbooking[C]. The 36th International Conference on Machine Learning, Long Beach, USA, 2019: 132–139.
- [14] QIU QinFu and CHEN Xiong. Behaviour-driven dynamic pricing modelling via hidden Markov model[J]. *International Journal of Bio-Inspired Computation*, 2018, 11(1): 27–33. doi: [10.1504/IJBIC.2018.090071](https://doi.org/10.1504/IJBIC.2018.090071).
- [15] LAWHEAD R J and GOSAVI A. A bounded actor-critic reinforcement learning algorithm applied to airline revenue management[J]. *Engineering Applications of Artificial Intelligence*, 2019, 82: 252–262. doi: [10.1016/j.engappai.2019.04.008](https://doi.org/10.1016/j.engappai.2019.04.008).
- [16] RAMASWAMY A and BHATNAGAR S. Stability of stochastic approximations with “controlled markov” noise and temporal difference learning[J]. *IEEE Transactions on Automatic Control*, 2019, 64(6): 2614–2620. doi: [10.1109/TAC.2018.2874687](https://doi.org/10.1109/TAC.2018.2874687).
- 卢 敏: 男, 1985年生, 副研究员, 博士, 主要研究方向为机器学习、强化学习.
- 张耀元: 女, 1996年生, 硕士生, 主要研究方向为民航收益管理、强化学习.
- 卢 春: 男, 1974年生, 高级工程师, 研究方向为航空公司收益管理.

责任编辑: 陈 倩