

5-15-2020

## Prediction of Flight Taxi-out Time in A Busy Airport Based on LWSVR

Zhiwei Xing

1. *College of Electronic Information and Automation, Civil Aviation University of China, Tianjin 300300, China;*

Songyue Jiang

1. *College of Electronic Information and Automation, Civil Aviation University of China, Tianjin 300300, China;*

Luo Qian

2. *The Second Research Institute of Civil Aviation Administration of China, Chengdu 610041, China;*

Luo Xiao

2. *The Second Research Institute of Civil Aviation Administration of China, Chengdu 610041, China;*

Follow this and additional works at: <https://dc-china-simulation.researchcommons.org/journal>



Part of the [Artificial Intelligence and Robotics Commons](#), [Computer Engineering Commons](#), [Numerical Analysis and Scientific Computing Commons](#), [Operations Research, Systems Engineering and Industrial Engineering Commons](#), and the [Systems Science Commons](#)

---

This Paper is brought to you for free and open access by Journal of System Simulation. It has been accepted for inclusion in Journal of System Simulation by an authorized editor of Journal of System Simulation.

---

## Prediction of Flight Taxi-out Time in A Busy Airport Based on LWSVR

### Abstract

**Abstract:** Aiming at improving the accuracy of predicting the flight taxi-out time in a busy airport, based on the local regression and weighted support vector regression, a prediction model of the locally weighted support vector regression is proposed. *The model uses the K nearest neighbor method to reduce the capacity of the training sample set and build a predictive model for each predicted sample. The bandwidth parameter of the Gaussian weighting function is optimized with the Mahalanobis distance between the forecast sample and training samples, and the weighting coefficients are obtained.* Combining the airport departure flight data in simulation analysis, the experimental results show that the accuracy of LWSVR within the error range is 83.33%, and the model is more stable.

### Keywords

taxi-out time, local regression, weighted support vector regression, KNN(K-Nearest Neighbor), Gaussian weighting function

### Recommended Citation

Xing Zhiwei, Jiang Songyue, Luo Qian, Luo Xiao. Prediction of Flight Taxi-out Time in A Busy Airport Based on LWSVR[J]. Journal of System Simulation, 2020, 32(5): 927-935.

# 基于 LWSVR 的繁忙机场航班滑出时间预测

邢志伟<sup>1</sup>, 姜松岳<sup>1</sup>, 罗谦<sup>2</sup>, 罗晓<sup>2</sup>

(1. 中国民航大学电子信息与自动化学院, 天津 300300; 2. 中国民用航空总局第二研究所, 四川 成都 610041)

**摘要:** 针对繁忙机场航班滑出时间预测准确率低的问题, 结合局部回归和加权支持向量回归, 提出基于局部加权支持向量回归的离港航班滑出时间预测模型。该模型采用  $K$  最近邻方法, 减小训练样本集容量, 并为每个预测样本构建一个预测模型。通过计算训练样本与预测样本间的马氏距离, 来优化加权支持向量回归中高斯核加权函数的带宽参数, 获得加权系数。结合某机场离港航班数据仿真分析, 实验结果表明模型在误差允许范围内的预测准确率达到 83.33%, 模型更加稳定。

**关键词:** 滑出时间; 局部回归; 加权支持向量回归;  $K$  最近邻; 高斯加权函数

中图分类号: U8; TP391.9 文献标识码: A 文章编号: 1004-731X (2020) 05-0927-09

DOI: 10.16182/j.issn1004731x.joss.18-0598

## Prediction of Flight Taxi-out Time in A Busy Airport Based on LWSVR

Xing Zhiwei<sup>1</sup>, Jiang Songyue<sup>1</sup>, Luo Qian<sup>2</sup>, Luo Xiao<sup>2</sup>

(1. College of Electronic Information and Automation, Civil Aviation University of China, Tianjin 300300, China;

2. The Second Research Institute of Civil Aviation Administration of China, Chengdu 610041, China)

**Abstract:** Aiming at improving the accuracy of predicting the flight taxi-out time in a busy airport, based on the local regression and weighted support vector regression, a prediction model of the locally weighted support vector regression is proposed. The model uses the  $K$  nearest neighbor method to reduce the capacity of the training sample set and build a predictive model for each predicted sample. The bandwidth parameter of the Gaussian weighting function is optimized with the Mahalanobis distance between the forecast sample and training samples, and the weighting coefficients are obtained. Combining the airport departure flight data in simulation analysis, the experimental results show that the accuracy of LWSVR within the error range is 83.33%, and the model is more stable.

**Keywords:** taxi-out time; local regression; weighted support vector regression; KNN(K-Nearest Neighbor); Gaussian weighting function

## 引言

提高机场运行效率现已经成为解决航班延误的重点研究领域之一, 其中航班滑出时间的准确预测是提高机场运行效率的关键一步。航班的滑出时

间是指飞机从撤轮档结束开始, 由停机位推出, 到实际起飞前的时间, 是反映机场运行效率的重要指标之一。目前, 多数机场均采用统一的默认滑出时间来预估离港航班的预计起飞时间, 对于繁忙机场, 由于受到航班起降数量、机位分配和机场布局等因素的影响, 实际滑出时间分布范围较大, 传统的默认滑出时间估计方法精度偏低。如果在航班推出前, 能准确预测航班的滑出时间, 不仅能够有效避免机场场面拥堵, 提高跑道利用效率, 并且能减轻飞机的燃油消耗, 减少航班延误。



收稿日期: 2018-09-07 修回日期: 2018-12-17;  
基金项目: 国家自然科学基金(U1533203);  
作者简介: 邢志伟(1970-), 男, 辽宁沈阳, 博士, 教授, 研究方向为民航装备与系统、机场交通信息与控制; 姜松岳(1990-), 男, 山东烟台, 硕士, 研究方向为机场交通信息与控制。

<http://www.china-simulation.com>

国内外对于航班滑出时间预测取得了一定的研究成果。Jordan 等<sup>[1]</sup>分析了影响滑行时间的相关因素,以滑行距离、航空公司、进离港航班数量等因素建立多元线性回归模型,但是此模型在反映机场运行复杂性方面存在不足。Ravizza 等<sup>[2]</sup>提出将飞机滑行路径中的转弯转角因素加入到滑行时间预测的多元线性回归模型中。此外, Balakrishna 等<sup>[3]</sup>使用基于非参数强化学习(RL)的方法,在随机动态规划的概率框架中提高滑行时间预测的准确性。Idris 等<sup>[4]</sup>将离港队列大小视为影响航班滑出时间的最关键因素,通过预测航班可能遇到的滑行冲突来估计离港队列大小。Lee 等<sup>[5]</sup>分别使用支持向量机、KNN 和神经网络模型等机器学习方法对滑出时间进行预测,但由于没有针对离港航班数据特点进行算法优化,所以导致预测偏差较大。

结合离港航班数据量大、分布范围广、周期性的特点,针对单跑道起飞、单跑道落地的跑道配置模式进行研究。结合局部回归思想和航班的周期性特点,从样本数量较大的训练集中,搜索出与每个预测样本的特征信息相似的训练样本,组成近邻训练集,为每个预测样本构建一个加权 SVR 预测模型。然后针对近邻训练集中训练样本与预测样本的相似度不同,利用带宽优化的高斯加权函数使每个训练样本获得不同的加权系数,从而获得不同的惩罚参数。通过结合局部回归与加权支持向量回归,构建滑出时间的局部加权支持向量回归(Locally Weighted Support Vector Regression, LWSVR)预测模型。

## 1 影响因素

以国内某机场的航班离港统计数据为依据,分析影响航班滑出时间的影响因素。在清除由于天气原因、飞机故障、军事活动等造成的航班未离港产生的数据后,为了提高预测的精度,将每天的运行数据依据时间分成每 15 min 一个时间段。选取关键的影响因素对于回归预测模型的建立至关重要,把尽可能多的因素考虑进去并不能提高预测精度,反而会导致较长时间的算法运行时间。

由于离港航班的滑行时间分布具有空间相关性和时间相关性,因此在分析影响因素时,前者考虑不同的滑行路径的影响,而后者则兼顾考虑时段内航班量的影响<sup>[6]</sup>。此外,人为因素也对滑出时间影响很大,因此将考虑不同航空公司的影响。具体分析如下。

### 1.1 滑行路径

不同机位的滑行路径通常情况下遵循就近原则,因此各机位的滑行路径基本固定。由于机场各机位距跑道距离不同,导致从各机位推出的航班的滑行距离不同,在其他样本特征相同的情况下,滑出时间与滑行距离成正比,定义滑行距离为  $D$ 。

在滑行过程中,飞机转弯前必须进行减速,从而增加了滑出时间,因此通常滑出时间会随转角数量增加而增加,定义转角数量为  $Z$ 。

### 1.2 时段内进离港航班量

通过分析发现,航班的滑出时间不仅与该时间段内推出和起飞的航班量有关,由于机场滑行道时间段内所能容纳的航班量有限,进港航班量也是影响滑出时间的重要因素。图 1 所示为某一机位在一天内 09:00-15:00 推出的飞机的滑出时间与进离港航班总量的关系。

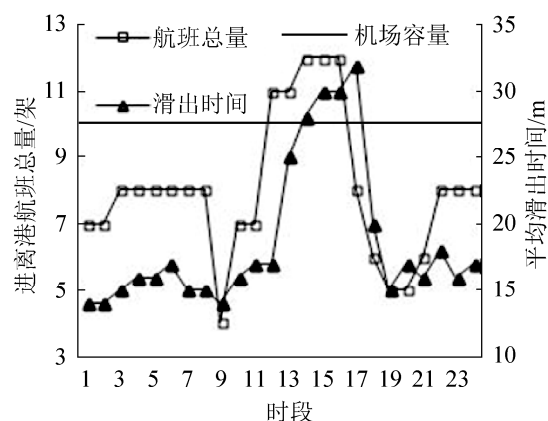


图 1 进离港航班总量与滑出时间对比曲线

Fig. 1 Comparison between incoming and outgoing flight volume and flight taxi-out time

时段内进离港航班量未达到该机场容量上限时,滑出时间受航班量波动影响不大,但当航班量

达到机场容量上限后, 此时跑道端头逐渐形成较长起飞等待队列, 因此平均滑出时间增长较快, 随着航班总量的减少, 起飞等待队列逐渐缩短, 因此平均滑出时间逐渐减小。

引入推出量( $P$ )、起飞量( $T$ )、到达量( $R$ ), 分别代表时间段内从停机位推出的航班数量, 实际起飞和计划到达的航班数量。其中, 由于滑行道容量的限制, 将到达量( $R$ )作为影响滑出时间的重要因素。

### 1.3 起飞等待队列

分析发现, 上一时段推出的航班很有可能在下一时段内起飞, 表明航班推出量对滑出时间的影响存在延时性, 因此引入滑行量( $N$ ), 指在时段结束时仍在场面滑行的飞机数量, 即场面上已经推出、未起飞的飞机总量, 我们定义滑行量( $N$ )与推出量( $P$ )、起飞量( $T$ )的关系式:

$$N(t_j) = N(t_{j-1}) + P(t_j) - T(t_j), j = 1, 2, \dots \quad (1)$$

式中:  $t_j$  为时段;  $j$  为时段编号。

在同一时段内, 在预测样本航班前推出的航班量为  $N^p$ , 在预测样本航班后计划推出的航班量为  $N_p$ 。在预测样本航班前推出的航班与前一时段推出但未起飞的航班组成预测样本航班的起飞等待队列。表达式如下:

$$Q(t_j) = N(t_{j-1}) + N^p(t_j) \quad (2)$$

图 2 给出了  $Q$  与滑出时间的线性相关性分析。

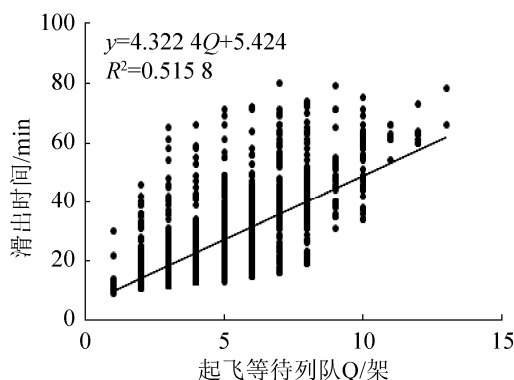


图 2 滑出时间与  $Q$  相关性  
Fig. 2 Correlation between taxi-out time and  $Q$

由图 2 可知,  $Q$  与滑出时间有较强的线性相关性,  $Q$  能较好的反映场面航班量的动态特性。

并且由于调度或其他紧急原因, 如某一航班在机位延误时间过长, 需优先起飞。此时计划后推出的航班可能需要超过当前预测的航班, 因此, 将  $N_p$  也作为输入样本的特征向量。

### 1.4 航空公司

由于各航空公司对于航班的滑出速度等限制不同, 并且当机场场面出现拥堵时, 存在航空公司之间的推出博弈等问题, 因此航空公司也影响着航班滑出时间。对于各个样本, 定义航空公司为一个二进制变量  $A_m, m \in [1, 39]$ , 其中  $m$  代表离港航班数据中 39 家不同航空公司。若某一样本属于  $A_1$  航空公司, 则  $A_1=1$ , 否则  $A_1=0$ 。

## 2 支持向量回归

支持向量回归在解决小样本、非线性系统回归拟合中表现出了巨大优势, 并且具有较强的泛化性。其基本思想是通过一个非线性映射  $\phi$ , 将数据  $x_i$  映射到高维特征空间  $H$ , 并在这个空间进行线性回归。具体的函数形式可以表示为:

$$f(x) = (\omega, \phi(x_i)) + b, \phi: R^n \rightarrow H, \omega \in R^n \quad (3)$$

式中:  $\omega$  为回归系数;  $b$  为偏置量;  $i$  为样本编号。

支持向量回归的目标函数为:

$$\begin{aligned} \min_{\omega, b, \xi_i, \xi_i^*} & \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) \\ \text{s.t.} & \begin{cases} y_i - \langle \omega, \phi(x_i) \rangle + b \leq \varepsilon + \xi_i^* \\ \langle \omega, \phi(x_i) \rangle + b - y_i \leq \varepsilon + \xi_i \\ \xi_i, \xi_i^* \geq 0 \end{cases} \end{aligned} \quad (4)$$

式中:  $C$  为惩罚参数;  $\varepsilon$  为误差要求;  $\xi_i$  和  $\xi_i^*$  为误差;  $y_i$  为预测样本实际值;  $N$  为训练样本数量; 惩罚参数  $C$  是权衡函数的平滑性和计算训练数据精度的常数。  $C$  表示对超出误差要求  $\varepsilon$  的样本数据的惩罚,  $C$  值越小意味着对超出误差要求的样本惩罚力度就越小, 从而增大训练误差。传统 SVR 模型中采用同一惩罚参数  $C$ , 但实际上对于训练集中不同的训练样本, 所需的惩罚力度是不同的, 传统模型显然无法满足这一要求。

引入拉格朗日乘子  $\alpha_i, \alpha_i^*$ , 并且根据 KKT 最优化

条件, SVR 的训练进程相当于解决二次凸规划问题:

$$\begin{aligned} \min_{\alpha, \alpha^*} & \frac{1}{2} \sum_{i,j=1}^N (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(x_i, x_j) + \dots \\ & \varepsilon \sum_{i=1}^N (\alpha_i + \alpha_i^*) - \sum_{i=1}^N y_i (\alpha_i - \alpha_i^*) \\ \text{s.t.} & \begin{cases} \sum_{i=1}^N (\alpha_i - \alpha_i^*) = 0 \\ 0 \leq \alpha_i, \alpha_i^* \leq C \end{cases} \end{aligned} \quad (5)$$

式中:  $K(x_i, x_j)$  为核函数。

本文使用高斯核函数, 其表达式为:

$$K(x_i, x) = \exp\left(-\frac{\|x_i - x\|^2}{2\sigma^2}\right) \quad (6)$$

最终, 得到的回归函数的表达式为:

$$\hat{f}(x) = \sum_{i=1}^N (\alpha_i - \alpha_i^*)K(x_i, x) + b \quad (7)$$

### 3 优化带宽的加权系数算法

#### 3.1 加权支持向量回归

在标准 SVR 模型中, 惩罚参数  $C$  是定值, 对于所有的训练样本, 其超出误差要求的惩罚是一样的。但实际数据中对于某个测试数据的预测, 不同的训练样本的贡献不同, 加权支持向量回归 (Weighted Support Vector Regression, WSVR) 的目的就是解决这一问题<sup>[7]</sup>。

在 WSVR 中, 每个训练样本会根据其与预测样本的相关性获得不同的权重, 相关性高的就获得较高的权重, 反之, 则获得较低的权重。在加权函数的选取中, 使用较为广泛的为高斯核加权函数:

$$W = e^{-\frac{d}{\sigma^2}} \quad (8)$$

在 WSVR 中, 优化后的风险函数变为:

$$\frac{1}{2} \|\omega\|^2 + \sum_{i=1}^N C_i (\xi_i + \xi_i^*) \quad (9)$$

式中:  $C_i$  为优化后的惩罚参数。

#### 3.2 加权系数

根据式(8), 得到如下加权函数:

$$W = e^{-\frac{M_f}{h}} \quad (10)$$

式中:  $M_f$  为预测样本与近邻训练样本的马氏距离;

$h$  为带宽参数。

式(10)中可以看出, 高斯核函数中不变的  $\sigma$  变成了可调的带宽参数  $h$ ,  $h$  是预测模型中很重要的参数。当训练样本与预测样本相似度最高时,  $W$  取得最大值, 反之, 则取得最小值。

带宽参数  $h$  表明了泛化范围的特征, 在模型中是一个很重要的参数。如果  $h$  无限大, LWSVR 模型会变成未加权的局部 SVR 模型, 如果  $h$  太小, 可能在提取的近邻训练集中得不到足够有效的支持向量<sup>[8]</sup>。

由于近邻训练集中, 每个样本对于预测的贡献度不同, 因此将  $h$  选择一个常数, 不能满足航班离港这样的非线性复杂系统的预测要求。本文通过计算马氏距离, 来优化带宽参数  $h$ , 从而得出加权系数, 以提高 LWSVR 模型的预测精度。马氏距离基于样本之间相关性, 可以解决欧氏距离中针对每个近邻训练集样本采用固有尺度和相关性。并且马氏距离与测量单位无关, 因此不会受特征值在数量级上的差异影响。

标准马氏距离公式为:

$$M = \sqrt{(\mathbf{X} - \mu)^T \mathbf{S}^{-1} (\mathbf{X} - \mu)} \quad (11)$$

式中:  $\mu$  为均值;  $\mathbf{S}^{-1}$  为逆协方差矩阵。

定义预测样本  $\mathbf{X}_f$  和训练集样本  $\mathbf{X}$  的马氏距离为:

$$M_f = \sqrt{(\mathbf{X} - \mathbf{X}_f)^T \mathbf{U}^{-1} (\mathbf{X} - \mathbf{X}_f)} \quad (12)$$

式中:  $\mathbf{X}$  属于预测样本  $\mathbf{X}_f$  的近邻训练集;  $\mathbf{U}^{-1}$  为每列删除列平均值后的逆协方差矩阵。

带宽参数  $h$  是  $M_f$  的函数:

$$h = g(M_f) = a \left( \frac{1 - bM_f}{M_f} \right)^2 + c \quad (13)$$

式中:  $a, b, c$  为常数。

通过应用边界条件<sup>[9]</sup>, 可以得到:

$$h = g(M_f) = (1 - \delta) \left( \frac{M_{\min} (M_{\min} - M_f)}{M_f (M_{\max} - M_{\min})} \right)^2 + \delta \quad (14)$$

式中:  $M_{\min}$  为  $h_{\max} = g(M_{\min}) = 1$  所对应的是近邻训

练集中的对预测样本最重要的样本;  $M_{\max}$  为  $h_{\min} = g(M_{\max}) = \delta$  所对应的是近邻训练集中的对预测样本最不重要的样本;  $\delta$  为一个接近于 0 的正常数, 取  $\delta=0.01$ 。

将式(12)代入式(8)中, 得到优化带宽  $h$  后的高斯核加权函数为:

$$W_i = e^{-\left(\frac{M_f}{(1-\delta)\left(\frac{M_{\min}(M_{\max}-M_f)^2}{M_f(M_{\max}-M_{\min})}\right) + \delta}\right)^2} \quad (15)$$

## 4 基于 LWSVR 的滑出时间预测

### 4.1 局部回归

局部回归基于某一预测样本的相邻样本是该预测该样本的最佳模型输入的原则, 使用训练数据的子集来训练该预测样本的局部拟合函数。这个子集通过计算输入空间中, 预测样本与训练样本的欧氏距离获得距离最近的  $k$  个样本, 并使用该子集训练获得 SVR 模型和相应参数<sup>[10]</sup>。第  $i$  个训练样本与预测样本的欧氏距离公式为:

$$d_i = \sqrt{\sum_{p=1}^N (V_p - V_{fp})^2} \quad (16)$$

式中:  $d_i$  为第  $i$  个训练集样本与预测样本间的欧氏距离;  $V_p$  为训练样本的第  $p$  个变量;  $V_{fp}$  为预测样本的第  $p$  个变量。

### 4.2 构建 LWSVR 模型

LWSVR 模型结合局部回归与 WSVR。在实际离港航班数据中, 不同的训练集样本对于预测的重要性不同, 所需的惩罚力度也不同。特征向量越接近预测样本的训练样本, 说明该训练样本中的离港航班在滑行路径、推出时的进离港航班量和起飞等待队列等特征向量值上更接近预测航班, 对预测的贡献越高, 相反, 则越低。因此, 对不同的近邻训练集样本, 计算不同的惩罚参数  $C$ 。优化后的惩罚参数为:

$$C_i = W_i \times C \quad (17)$$

式中:  $W_i$  为第  $i$  个训练样本的加权系数。

step 1: 对航班离港数据进行预处理, 提出关键特征向量信息, 并将组成的样本集分为训练集和测试集;

step 2: 寻找当前预测样本近邻训练集, 并计算模型参数  $C$  和  $\epsilon$ ;

step 3: 计算带宽参数  $h$  和高斯加权函数值  $W_i$ ;

step 4: 利用上一步所得参数重新计算模型参数  $C_i$ ;

step 5: 使用近邻训练集和新的模型参数  $C_i$ 、 $\epsilon$  训练 LWSVR 模型, 并计算预测值; 判断是否为最后一个预测样本, 不是则返回 step 2, 是则结束计算, 输出预测结果。

图 3 所示为 LWSVR 预测模型的计算过程。

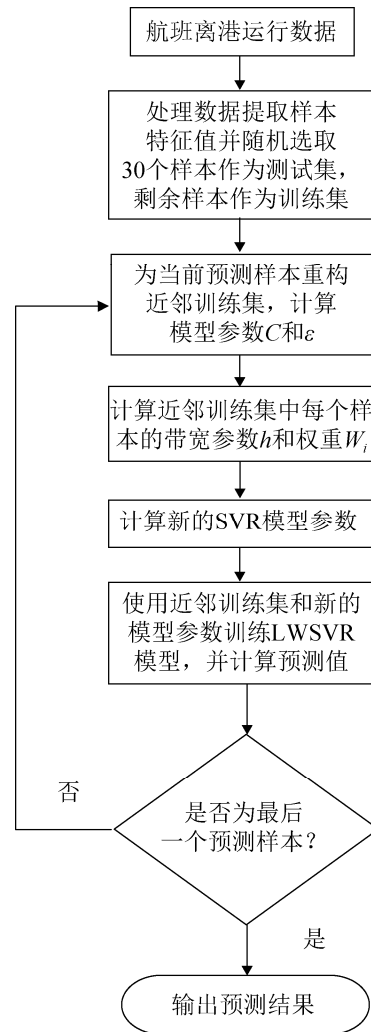


图 3 LWSVR 模型流程图  
Fig. 3 Flowchart of LWSVR model

## 5 实验分析

### 5.1 数据预处理

实验数据为国内某机场 2016 年 5 月 1 日至 5 月 31 日的航班离港运行数据, 数据中具体包含航班号、实到时间、实飞时间、停机位、撤轮档结束时间等信息。具体处理步骤如下:

(1) 在清除由于天气、飞机故障、军事活动等原因造成的航班未离港所产生的数据后, 为了提高预测的精度, 将每天的样本数据依据时间分成每 15 min 一个时间段。

(2) 提取航班号、执行日期、实到、实飞时间、停机位、撤轮档结束时间等关键数据, 经过计算, 转换成航空公司( $A_m$ )、到达量( $R$ )、推出量( $P$ )、滑

行距离( $D$ )、转角数量( $Z$ )、滑行量( $N$ )时段内的  $N^p$  和  $N_p$ 。根据式(2)计算得出每个待预测航班推出时的起飞等待队列( $Q$ )。最终每个样本包含的特征变量为  $\mathbf{X} = \{D, Z, N_p, R, Q, A_m\}$ ,  $m = 1, 2, \dots$ 。

(3) 对于每个样本, 因为不同特征向量的数量级和量纲不同, 直接计算欧式距离寻找预测样本的近邻训练集时, 数量级较小的特征向量在预测模型中起的作用会受影响。例如  $D$  的数量级远大于其他特征向量的数量级, 如果直接计算会影响除  $D$  外的其他特征向量在预测模型中的作用。因此在计算欧式距离前, 先对样本数据进行归一化处理, 使所有数据映射到 $[-1, 1]$ 的区间, 消除量纲的不利影响。

经过上述步骤预处理得到的部分样本数据如表 1 所示。

表 1 预处理后的输入样本  
Tab. 1 Preprocessed input samples

样本	滑行距离( $D$ )	转角数量( $Z$ )	后推出航班量( $N_p$ )	到达量( $R$ )	等待队列长度( $Q$ )	$A_1$	...	$A_{39}$
$X_1$	0.738 402	1.000	-0.111 111	0.714 285	-1.000 0	1	...	-1
$X_2$	0.944 331	1.000	-0.333 333	-0.714 285	-0.666 7	-1	...	-1
$X_3$	0.294 410	0.600	-0.555 555	-0.714 285	-0.333 3	-1	...	-1
$X_4$	-1.000 000	-1.000	0.333 333	-0.714 285	0.000 0	-1	...	-1
$X_5$	0.976 465	1.000	-0.777 777	-0.428 571	0.333 3	-1	...	1

### 5.2 参数选择

为得到准确有效的 LWSVR 模型, 模型参数的正确选择至关重要。

(1) 近邻值  $k$  的选取, 表征了近邻训练集的容量<sup>[11-12]</sup>。 $k$  值的选取很大程度上与训练集的具体情况和样本的具体构成有关。一般情况下, 预测精度首先会随着  $k$  值的增加而逐渐增加, 随后趋于稳定, 然后随着  $k$  值的继续增加预测精度会下降。本文运用交叉验证法, 采用对预测值的平均绝对误差百分比的均值来确定最优  $k$  值。由于样本的维数为 44, 马氏距离计算要求样本数大于维数, 否则协方差矩阵  $S$  不可逆, 即要求近邻训练集容量  $k > 44$ 。取  $k_{\min} = 45$ ,  $k_{\max} = 100$ , 经过计算, 如图 4 可知,  $k$  值取 52 时预测值的平均绝对误差百分比最小。

(2) 在 SVR 模型中, 惩罚参数  $C$  和损失函数  $\varepsilon$

对预测精度影响很大, 本文采用基于交叉验证的网格搜索法<sup>[13]</sup>, 寻找针对近邻训练集的最优  $C$  和  $\varepsilon$ 。即对  $C$  和  $\varepsilon$  分别取 E 和 F 个值, 对  $E \times F$  个  $(C, \varepsilon)$  的组合, 分别训练不同的 SVR 模型, 并将均方误差 MSE 作为评价指标<sup>[14]</sup>。获得 MSE 最小的一组  $(C, \varepsilon)$  即为最优的  $C$  和  $\varepsilon$ 。

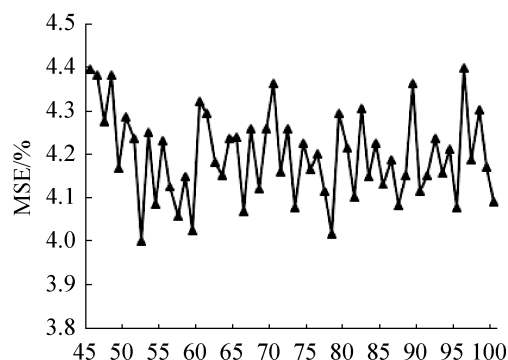


图 4 最优  $k$  值选择结果

Fig. 4 Optimal  $k$  value selection result



### 5.3 预测模型及结果分析

从预处理后的实验数据样本中随机选取 30 个样本作为测试数据, 剩余的作为训练集数据。对于预测结果的误差评价, 使用平均绝对误差(MAE), 平均绝对百分误差(MAPE)来衡量模型的预测效果。为了说明提出模型的有效性, 分别使用 BP 神经网络、标准 SVR 模型和局部 SVR 模型与之进行对比分析。神经网络是一种预测航班滑行时间的常用方法<sup>[15-16]</sup>, 而 BP 神经网络又是目前流行的一种神经网络模型。BP 神经网络使用归一化后的样本

数据, 输入节点数为 44, 输出节点数为 1, 隐含层节点个数为 10; 标准 SVR 使用所有训练集样本训练 SVR 模型, 且使用同一惩罚参数 C 训练预测模型; 局部 SVR 使用近邻训练集训练 SVR 模型, 为每个预测样本建立 SVR 模型, 并分别进行预测, 但训练预测模型时仍然使用同一惩罚参数 C。第 4 种方法使用本文提出 LWSVR 方法训练预测模型。

图 5 所示为 4 种方法的预测结果对比。表 2 为 4 种预测模型的预测准确率和误差指标。

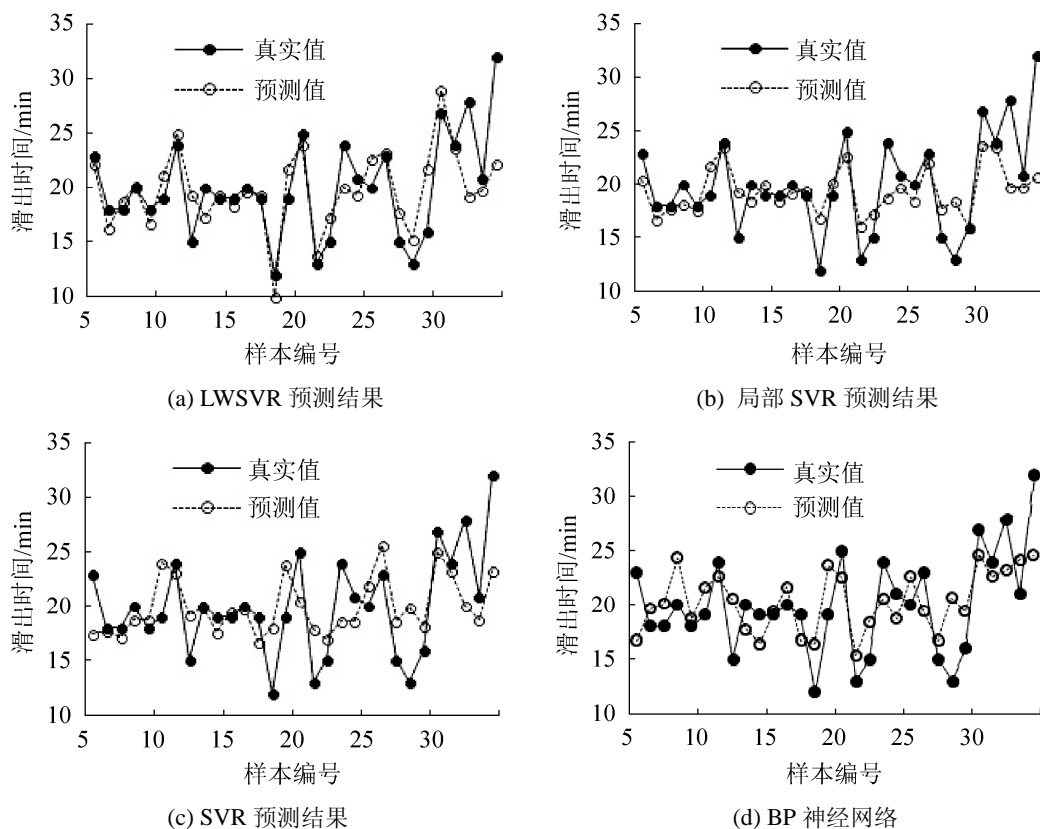


图 5 预测结果对比

Fig. 5 Comparisons of predictive

表 2 预测准确率、MAE、MAPE 对比

Tab. 2 Comparisons of predictive accuracy and MAE and MAPE

指标	LWSVR	局部 SVR	SVR	BP
准确率( $\pm 3$ min)/%	83.33	73.33	60.00	56.67
MAE/min	1.233	2.148	3.059	3.167
MAPE/%	6.19	10.57	16.23	16.85

BP 神经网络预测结果的准确率低于 SVR 模型, 并且 MAE 和 MAPE 大于 SVM 模型, 因此预测稳定性不如 SVR 模型。并且 BP 神经网络容易陷入局部最优解。

SVR 模型在所有训练集样本中寻找支持向量, 所获的支持向量个数远大于 LSVR 和 LWSVR 模

型。但在实际离港航班数据中,与预测样本在机位、推出时的进离港航班量、等待起飞队列长度等特征向量中有一个或多个相近的航班数据才有助于准确预测。训练样本集中存在很多样本特征与预测样本特征差异较大的训练样本,因此这类回归拟合所得的函数对于单个航班预测来说偏差较大,而对于实际需要,往往对单个航班的滑出时间预测的需求比较大,并且使用数量较大的支持向量训练预测模型相对来说耗时较长。

对于局部 SVR 模型,缩小了针对单个预测样本的训练集容量,得到近邻训练集,减少了模型训练时间,但是局部 SVR 模型没有解决 SVR 模型中对于不平衡数据的惩罚需求不同的问题。

LWSVR 模型使用局部 SVR 与 WSVR 结合的方法,针对每个预测样本寻找一个近邻训练集,来找到该预测样本的局部拟合函数,相对于标准 SVR 缩短了训练时间;由于近邻训练集中每个训练样本对预测的贡献度不同,为每个样本赋予不同的惩罚参数  $C$ ,训练所得的回归拟合函数较局部 SVR 进一步提高了预测精度。LWSVR 的误差范围内预测准确率达到 83.33%,明显高于其他模型。

LWSVR 模型的 MAE 和 MAPE 明显比其他 3 种模型的低,说明 LWSVR 预测模型具有更小的误差波动,预测模型更加稳定。总的来说,LWSVR 模型的预测性能明显优于其他 3 种模型。

## 6 结论

(1) 提出的航班滑出时间预测模型满足了实际航班离港运行中,对于单个航班在不同机位、场面航班量等可变条件下的滑出时间预测的要求。

(2) 预测模型训练时间短,准确率更高,不需要计算上一时段的平均航班滑出时间,易于实现。为后续根据预测的滑出时间制定推出策略提供了重要依据。

## 参考文献:

[1] Jordan R, Ishutkina M A, Reynolds T G. A statistical

learning approach to the modeling of aircraft taxi time[C]. Digital Avionics Systems Conference. Salt Lake City: IEEE, 2010.

[2] Ravizza S, Atkin J A D, Maathuis M H, et al. A combined statistical approach and ground movement model for improving taxi time estimations at airports[J]. Journal of the Operational Research Society (S0160-5682), 2013, 64(9): 1347-1360.

[3] Balakrishn A P, Ganesan R, Sherry L. Accuracy of reinforcement learning algorithms for predicting aircraft taxi-out times: A case-study of Tampa Bay departures[J]. Transportation Research Part C Emerging Technologies (S0968-090X), 2010, 18(6): 950-962.

[4] Idris H, Clarke J P, Bhuvra R, et al. Queuing Model for Taxi-Out Time Estimation[J]. Air Traffic Control Quarterly (S1064-3818), 2001, 10(1): 1-22.

[5] Lee H, Malik W, Jung Y C. Taxi-out time prediction for departures at charlotte airport using machine learning techniques[C]. Aiaa Aviation Technology, Integration, and Operations Conference. Washington, D.C: AIAA, 2016: 3910.

[6] 刘丽华. 市场机制下飞机推出时隙分配模型与算法研究[D]. 哈尔滨: 哈尔滨工业大学, 2017.

Liu Lihua. Research on model and algorithm for aircraft pushback slot allocation under market mechanism[D]. Harbin: Harbin Institute of Technology, 2017.

[7] Zeng J S, Gao C H, Luo S H. Identification of LPV system using locally weighted technique[J]. Applied Mathematics: A Journal of Chinese Universities (S1005-1031), 2010, 25(4): 411-419.

[8] Wang H, Cao C, Leung H. An improved locally weighted regression for a converter re-vanadium prediction modeling[C]. World Congress on Intelligent Control and Automation. Dalian: IEEE, 2006: 1515-1519.

[9] 石柯, 陈洪生, 张仁同. 一种基于支持向量回归的 802.11 无线室内定位方法[J]. 软件学报, 2014, 25(11): 2636-2651.

Shi Ke, Chen Hongsheng, Zhang Rentong. Indoorlocation method based on support vector regression in 802.11 wireless environments[J]. Journal of Software, 2014, 25(11): 2636-2651.

[10] 王翔, 陈小鸿, 杨祥妹. 基于 K 最近邻算法的高速公路短时行程时间预测[J]. 中国公路学报, 2015, 28(1): 102-111.

Wang Xiang, Chen Xiaohong, Yang Xiangmei. Short term prediction of expressway travel time based on KNN algorithm[J]. China Journal of Highway and Transport,

- 2015, 28(1): 102-111.
- [11] Wang Z Y, Peng Z R. A PSO-SVM model for short-term travel time prediction based on bluetooth technology[J]. Journal of Harbin Institute of Technology (S1005-9113), 2015, 22(3): 7-14.
- [12] Ravizza S, Chen J, Atkin J A D, et al. Aircraft taxi time prediction: comparisons and insights[J]. Applied Soft Computing (S1568-4946), 2014, 14(1): 397-406.
- [13] 张洪海, 杨磊, 别翌荟, 等. 终端区进场交通流广义跟驰行为与复杂相变分析[J]. 航空学报, 2015, 36(3): 949-961.
- Zhang Honghai, Yang Lei, Bie Yihui, et al. Analysis on generalized following behavior and complex phasetransition law of approaching traffic flow in terminal airspace[J]. Acta Aeronautica Et Astronautica Sinica, 2015, 36(3): 949-961.
- [14] 戴文雯, 别翌荟, 张洪海, 等. 基于 NetLogo 的终端区交通流仿真[J]. 航空计算技术, 2014, 44(1): 30-33.
- Dai Wenwen, Bie Yihui, Zhang Honghai, et al. Simulation of Air Traffic Flow Based on Net Logo[J]. Aeronautical Computing Technique, 2014, 44(1): 30-33.
- [15] 谷润平, 崔朋, 唐建勋, 等. 基于 D\*算法的场面滑行动态规划研究[J]. 科学技术与工程, 2015, 15(1): 315-319.
- Gu Runping, Cui Peng, Tang Jianxun, et al. Research on Dynamic Planning of Taxi Path Based on D\* Algorithm[J]. Science Technology and Engineering, 2015, 15(1): 315-319.
- [16] Chatterji G, Zheng Y. Wheels-off time prediction using surface traffic metrics[C]. Aiaa Aviation Technology, Integration, & Operations. Indianapolis: AIAA, 2012: 5699.