

基于非参数核密度估计的集装箱码头交通需求预测模型

马梦知, 范厚明*, 黄莒森, 孔 靓, 岳丽君

(大连海事大学 交通运输工程学院, 辽宁 大连 116026)

摘要: 针对现有集装箱集疏港时间概率分布建模中需要假设参数分布的问题, 基于非参数核密度估计理论建立集装箱码头交通需求预测模型。非参数核密度估计的核函数选取高斯核, 最优带宽由交叉验证法求得, 通过检验、K-S 检验和后验检验对比分析了核密度估计与两种传统参数模型的估计效果, 并应用该模型预测 DCT 码头的交通需求。结果表明: 非参数核密度估计模型具有更高的拟合精度、稳定性和适用性, 得到的概率密度曲线能更加准确反映出集装箱集港时间和进口箱疏港时间的整体分布形态。基于非参数核密度估计的集装箱码头交通需求预测模型具有比传统的参数模型更高的预测精度, 可为集装箱码头基础设施规划、集疏港通道的道路交通管理、码头资源优化配置和调度等问题的研究提供更准确的交通量和作业任务量预测。

关键词: 集装箱码头; 交通需求预测; 非参数核密度估计; 交叉验证法

中图分类号: U169 **文献标志码:** A

Traffic demand forecasting model for container terminal based on non-parametric kernel density estimation

MA Meng-zhi, FAN Hou-ming*, HUANG Ju-sen,
KONG Liang, YUE Li-jun

(Transportation Engineering College,

Dalian Maritime University, Dalian 116026, China)

Abstract: To address the inadequacies associated with present parametric density estimations for containers' delivery and pick up time distributions, a traffic demand forecasting model based on non-parametric kernel density estimation was devel-

oped. Gaussian kernel was chosen as the kernel function and the optimal window width was obtained by the cross-validation method. The χ^2 test, K-S test and posteriori test were used to compare the goodness-of-fit of the proposed probabilistic model and two conventional parametric distribution models, and the proposed model was applied to forecast the traffic demand of DCT. The results demonstrate that the proposed non-parametric estimation has better accuracy, stability and applicability. The probability density curve obtained can more accurately reflect the overall distribution pattern of containers' delivery and pick up time. The traffic demand forecasting model based on non-parametric kernel density estimation has higher prediction accuracy than that of the conventional parametric distribution models, which can provide more accurate traffic volume and task volume prediction for the infrastructure planning of container terminal, road traffic management, allocation and scheduling of terminal resources.

Key words: container terminal; traffic demand forecasting; non-parametric kernel density estimation; cross-validation

0 引言

在集装箱船舶大型化的趋势下, 集装箱船舶装载量普遍提升, 一艘大型集装箱船的集装箱装载量远超以往, 而且, 由于大型集装箱船对港口的硬件设施要求很高, 大型集装箱船只能集中挂靠在少数几个有能力接纳大型船舶的港口, 使得大批货流集中在少数的枢纽港, 给这些港口的集疏港作业带来空前的压力。同时, 由于船舶、集装箱、

收稿日期: 2018-12-21; 修回日期: 2019-01-14.

基金项目: 国家自然科学基金资助项目(61473053); 辽宁省重点研发计划指导计划项目(2018401002).

作者简介: 马梦知(1989-), 女, 博士生, E-mail: mengzhi1440@126.com; 范厚明*(1962-), 男, 博士, 教授, 博士生导师, E-mail: fhm468@163.com.

集疏运车辆到港的不平衡性,造成了集装箱码头忙闲不均,直接影响了港口的生产效率。因此,迫切需要更科学的方法刻画集装箱集疏港时间的分布规律,准确预测集疏港交通需求,以提高港口各个生产环节有效计划、组织、调度与控制水平,解决集装箱码头忙闲不均的问题。

目前,国内外学者较少关注集装箱集疏港时间的问题,而是更多关注于对集疏港交通流的研究。在集装箱集疏港时间的预测方面,Kourounioti等^[1-2]先后使用泊松回归模型和人工神经网络模型对进口箱的在港停留时间的分布规律进行了研究,但该两个模型均为离散模型,且均是以天为最小时间单位的,所得结果无法满足集疏港道路交通管理、码头资源优化配置和调度的需要。针对集疏港交通流的研究基本可分为两类:一类沿用城市道路交通需求预测方法对港口交通需求进行预测;另一类利用统计学原理以及模型仿真等方法对港口交通流的分布特征进行分析。在港口交通需求预测方面,杨龙海等^[3]根据深圳盐田港保税区的生产作业能力,利用调整后的四阶段法对盐田保税港区的货运车辆的交通需求进行预测;Attila等^[4]根据一周内每天每小时的平均车辆到达率对港区交通需求进行预测。滕藤等^[5]利用支持向量机和卡尔曼滤波相结合方法对集卡到达时间进行预测。张莎莎等^[6]基于船期表利用季节性神经网络智能预判模型对天津港集装箱码头闸口的交通需求进行预测。虽然考虑了船期,但是每个班次分配的集疏港时长均是6天,显然与码头的实际生产情况不符。在港口交通流的分布特征方面,陈家源等^[7]利用统计学原理对到港、离港的船舶流、车流以及货流的特征与规律进行研究,但是,与文献[2-5]一样,均忽略了港区集疏运交通的派生性,即港口道路交通是由船舶装卸作业派生来的。船舶到离港情况对港区集疏运交通的影响很大,如果对港口交通需求进行分析预测时忽略了船舶的到离港情况,就会使分析结果存在一定偏差。杨忠振等^[8-9]依据船舶对集疏港车辆进行分组统计,利用概率分布拟合方法,分别对集装箱码头、散货码头、杂货码头生产作业诱发的交通流发生时间的概率分布进行分析。郭子坚等^[10]利用计算机仿真技术对集装箱港区货运交通的随机时变特性进行了分析。

对前人研究成果的梳理发现,针对集装箱集疏港时间的分布规律的研究较少,同时,文献[5]和文献[6]沿用道路交通需求预测方法进行集疏港交通流预测时,又需要依赖一些实时信息。实时更新的交通流信息虽然便于道路交通管理,但往往在集疏港作业开始前就需对集装箱码头资源进行优化配置和调度,此时码头能够掌握的信息一般只有船舶预计到离港时间和装卸箱量等信息,要想仅利用这些信息对作业任务量进行准确预测,概率分布拟合方法更为合适。但是,现有研究利用概率分布拟合方法研究集疏港交通流时间分布规律时,使用的均是参数统计的方法,即假设其服从贝塔分布、韦伯分布等已知的经典分布,再对假定的分布模型的未知参数进行估计和检验,但现实并非如此,很多集装箱码头的集疏港交通流时间的分布与这些已知的经典分布均有较大差距。传统的参数统计方法得到的集疏港交通流时间的分布可能会背离实际,不具有普适性。此时,就需要突破传统参数统计的框架结构,寻找新的统计学方法。统计学的另一个分支学科——非参数统计,就能实现在不假定总体分布服从确定分布的情况下,仅依靠从数据本身获得的信息对总体分布进行统计推断。正是因为非参数统计方法对总体的假定较少,因此,具有很好的稳定性和广泛的适用性。为此,本文提出基于非参数核密度估计的集装箱码头交通需求预测模型,在对DCT集装箱码头进口箱的疏港时间以及出口箱的集港时间分布规律进行研究的基础上,对集装箱码头的吸引交通量进行预测。

1 数据收集与分析

集装箱码头闸口是集装箱进出码头的出入口,因此,将闸口作为集装箱集疏港信息采集的观测点。从大连集装箱码头有限公司(DCT)TOP+系统中采集连续2周时间内94艘次集装箱船的13 906个出口箱和10 322个进口箱的属性信息,这些信息包括送/提箱集卡的进港时刻以及对应的车牌号、装载状态、船舶、航次、船舶集疏港时间等信息。由于每个艘次船舶集/疏港的开始时刻以及时间窗的长度均不同,要对所有船舶出口箱的集港时间在船舶集港时间窗内的分布以及进口箱疏港时间在船舶疏港时间窗内的分布进行整体分

析,就需要先对采集到的出口箱/进口箱的集/疏港数据进行标准化处理:

$$t_{ij}^{E'} = 100(t_{ij}^E - T_j^{ES}) / (T_j^{EE} - T_j^{ES}) \quad (1)$$

$$t_{ij}^{I'} = 100(t_{ij}^I - T_j^{IS}) / (T_j^{IE} - T_j^{IS}) \quad (2)$$

式中: t_{ij}^E 表示船舶 j 的第 i 个出口箱的集港时刻,即被集卡送到码头闸口的时刻; t_{ij}^I 表示船舶 j 的第 i 个进口箱的疏港时刻,即提箱集卡到达码头闸口的时刻; T_j^{ES} 和 T_j^{EE} 分别表示船舶 j 的集港时间窗的开始时刻和结束时刻; $T_j^{IE} - T_j^{ES}$ 表示船舶 j 的集港时间; T_j^{IS} 和 T_j^{IE} 分别表示船舶 j 的疏港时间窗的开始时刻和结束时刻; $T_j^{IE} - T_j^{IS}$ 表示船舶 j 的疏港时间; $t_{ij}^{E'}$ 和 $t_{ij}^{I'}$ 分别表示标准化后的集港时刻和疏港时刻。

根据标准化后的集港时刻和疏港时刻分别绘制进口箱疏港时间和出口箱集港时间分布直方图,如图 1 和图 2 所示。可以看出,出口箱集港时间与进口箱疏港时间的分布规律存在较大差别,因此,需要分别进行研究。通过直方图可以看出样本所属的总体是不对称的,且存在多个尖峰(如在临近截止时间时,集/疏港箱量均会再次出现一个小高峰)。目前被广泛使用的贝塔分布、韦伯分布等已知的经典分布很难准确描述总体的这些分布特征,而且,不同港口的集港/疏港时间的分布规律也会有明显差别,依据某一经典分布假设的传统概率分布模型明显不具有普适性和通用性。相比之下,非参数统计方法对总体的假定较少,更适合分析集装箱集/疏港时间的分布规律。

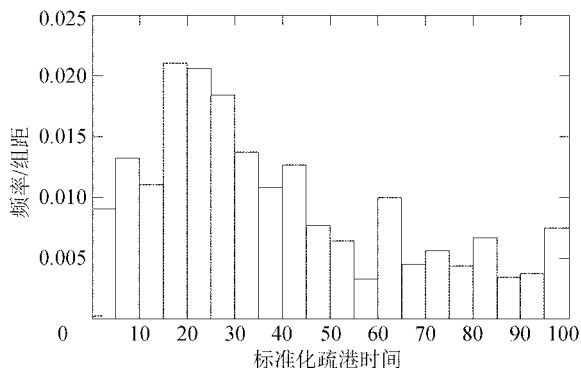


图1 进口集装箱疏港时刻的分布直方图

Fig.1 Distribution histogram of import containers' pick up time

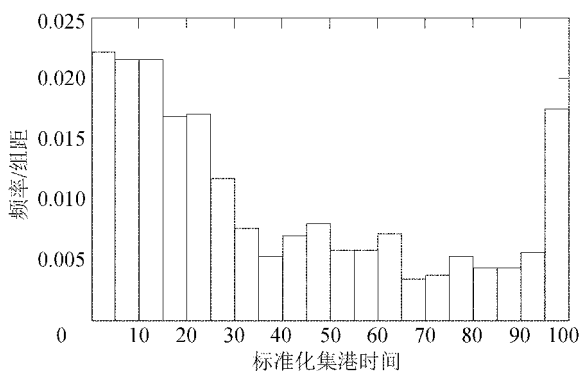


图2 出口集装箱集港时间的分布直方图

Fig.2 Distribution histogram of export containers' delivery time

2 集疏港时间分布的非参数估计

非参数统计方法这种统计推断方法不需要提前假定总体服从某一确定的分布族,因此,适用于任何总体分布信息未知或虽已知但不能用有限个参数刻画的统计问题,即仅根据样本数据直接对总体的分布形式进行统计推断^[11]。因此,非参数统计方法具有广泛的适用性。同时,非参数统计方法能够有效避免参数统计因对总体分布假设不当导致统计推断结果错误的问题,具有较好的稳定性。但是,在总体分布族已知情况下,非参数统计方法的效率要低于参数统计方法^[12]。

2.1 非参数核密度估计模型

假设 X_1, X_2, \dots, X_n 为标准化的集装箱的集港/疏港时刻的 n 个样本,集港/疏港时间的概率密度函数为 $f(x)$, 定义域为 $[a, b]$, 则非参数核密度估计的公式为^[13]:

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n k((x - X_i)/h) \quad (3)$$

式中: n 为样本容量; h 为带宽; $k(\cdot)$ 为核函数。

常用的核函数有: Triangle 核函数、Epanechnikov 核函数、Quartic 核函数、Triweight 核函数、Gaussian 核函数以及 Cosine 核函数等。研究证明,当带宽确定时,核函数的选择对于估计结果的影响较小^[11-12],因此,在核函数的选择上不需要进行过多讨论。高斯核函数具有良好的光滑度和可微性^[14],因此,本文选取高斯核函数进行非参数核密度估计,即

$$k(v) = e^{-v^2/2} / \sqrt{2\pi} \quad (4)$$

2.2 带宽的选择

带宽 h 又称光滑参数,其取值对核估计结果的影响较大,直接关系到核度估计能否成功。但是,现有利用统计学原理对于集疏港交通流时间分布规律进行研究的文献,并未对带宽的选择进行讨论,仅凭经验给出,不同的样本数量、总体分布使用的却是相同的带宽。图 3 为出口箱集港时间的核密度估计随带宽 h 的变化曲线。由图 3 可以看出,当带宽分别取 $h = 0.1$ 、 $h = 5$ 、 $h = 10$ 时,出口箱集港时间的核密度估计偏差随带宽增大而增大,即带宽越小,核估计的偏差越小、拟合程度越高,但曲线的光滑性越差,呈多峰不稳定的特征。随着带宽增大,曲线的光滑性虽然得到提高,但是一些重要峰值被掩盖,核估计的偏差增大、拟合程度降低。带宽的选择方法主要有经验法则法和交叉验证法。

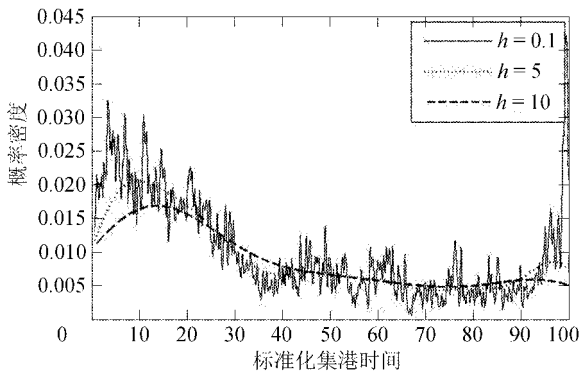


图 3 出口箱集港时间的核密度估计随带宽 h 的变化曲线

Fig. 3 The density estimation curve of export containers' delivery time with different bandwidth

(1) 经验法则法

通常使用 Rosenblatt 提出的积分均方误差 (MISE) 来评判核密度估计的精度。MISE 可以表示为:

$$\begin{aligned} \text{MISE}(\hat{f}) &= \int_{-\infty}^{\infty} E[\hat{f}(x) - f(x)]^2 dx = \\ &= c_{02}/nh + \frac{1}{4} \int [f''(x)]^2 dx c_{21}^2 h^4 + o(1/nh) + o(h^4) \end{aligned} \quad (5)$$

式中: $c_{02} = \int k^2(v) dv$; $c_{21} = \int v^2 k(v) dv$.

忽略 $\hat{f}(x)$ 的 MISE 高阶项,只考虑前两个主项,即渐近积分均方误差 (AMISE),使得 AMISE 最小的带宽即为最优带宽 h_0 :

$$h_0 = \left[c_{02} / \int [f''(x)]^2 dx c_{21}^2 \right]^{1/5} n^{-1/5} \quad (6)$$

但是,由于概率密度函数 $f(x)$ 是未知的,式 (6) 中的 $f''(x)$ 也是未知的。在这种情况下,Silverman 提出的经验法则法 (Rule-of-Thumb, ROT) 被广泛应用于核密度估计的带宽选择。假定 $f(x)$ 属于方差为 σ^2 的正态分布族,并选取 Gauss 核,得到的最优带宽为:

$$h_0 = 1.06 \hat{\sigma} n^{-1/5} \quad (7)$$

式中: $\hat{\sigma} = \sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 / (n-1)}$; \bar{X} 为样本均值。

(2) 交叉验证法

从 ROT 求取带宽的过程不难发现,其最大的问题是要事先假定 $f(x)$ 属于某一分布族,而核密度估计最大的特点就是仅依靠从数据本身获得的信息对总体分布进行统计推断,提前并不知道分布函数的形式信息。如果总体分布与假定的分布族相差甚远,经验法则法求出的带宽会直接影响估计的精度。实际应用中,经验法则法仅在求取单峰、比较对称、尾部概率不是很大的总体分布的最优带宽时效果比较好,但从图 1 和图 2 可以发现,集装箱的集疏港时间分布均存在多峰、不对称的现象。交叉验证法正好弥补了 ROT 的不足,该方法能够直接由数据“自动”产生最优带宽。

Rudemo^[15] 和 Bowman^[16] 最早提出了用交叉验证法 (Cross-Validation, CV) 确定最优带宽,该方法将使得积分平方误差 (ISE) 达到最小的带宽被视为最优带宽,ISE 可表示为:

$$\begin{aligned} \text{ISE}(\hat{f}) &= \int_{-\infty}^{\infty} [\hat{f}(x) - f(x)]^2 dx = \\ &= \int \hat{f}^2(x) dx - 2 \int \hat{f}(x) f(x) dx + \int f^2(x) dx \end{aligned} \quad (8)$$

由于 $\int f^2(x) dx$ 与带宽的选择无关,因此,最优带宽等价于最小化 $\int \hat{f}^2(x) dx - 2 \int \hat{f}(x) f(x) dx$ 的带宽。令 $I_1 = \int \hat{f}^2(x) dx$, $I_2 = 2 \int \hat{f}(x) f(x) dx$, 则有:

$$I_1 = \sum_{i=1}^n \sum_{j=1}^n k^* \left((X_i - X_j) / h \right) / (n^2 h) \quad (9)$$

式中: $k^*(v) = \int k(v+t)k(t)dt$.

由于 $\int \hat{f}(x)f(x)dx = E[\hat{f}(x)]$, 可得 I_2 的无偏估计为:

$$\hat{I}_2 = \sum_{i=1}^n \hat{f}_{-i}(X_i) / n = \sum_{i=1}^n \sum_{j=1, j \neq i}^n k((X_i - X_j)/h) / (n(n-1)h) \quad (10)$$

式中: $\hat{f}_{-i}(X_i)$ 表示将第 i 个观察点剔除后的核密度估计量.

因此, 交叉验证法所求的最优带宽 $h_v(\hat{f})$ 可表示为:

$$h_v(\hat{f}) = \arg \min_{h>0} \{I_1 - 2\hat{I}_2\} \quad (11)$$

2.3 分布拟合效果检验标准

(1) χ^2 检验

由于集装箱集疏港时间的总体分布是未知的, 这种情况下, 最常用的分布拟合检验方法为 χ^2 拟合优度检验. 将样本数据 X_1, X_2, \dots, X_n 分成不相交的 k 个类别. 由于已经对采集到的出口箱/进口箱的集/疏港时刻进行了标准化处理, 因此, 本文将数据分为 $k=10$ 类, 各类的观测频数如表 1 所示. χ^2 统计量的计算公式为:

$$\chi^2 = \sum_{i=1}^k (N_i - e_i)^2 / e_i \quad (12)$$

式中: N_i 为第 i 个类别里的样本观测频数; e_i 为第 i 个类别的期望频数, $e_i = n[\hat{F}(C_i) - \hat{F}(C_{i-1})]$; $\hat{F}(x)$ 为累积分布函数的核估计.

$$\hat{F}(x) = \sum_{i=1}^n K((x - X_i)/h) / n \quad (13)$$

式中: $K(u) = \int_{-\infty}^u k(v)dv$.

当 χ^2 大于临界值 $\chi_{k-m-1}^2(1-\alpha)$ 时, 表明在显著性水平 α 下, 拒绝对集装箱集疏港时间分布模型的假设. 即核估计方法得到的密度函数不是集装箱集疏港时间总体的密度函数. m 表示分布模型需要估计的参数数目, 由于核密度估计表达式中的带宽 h 是需要估计的参数, 令 $m=1$.

(2) K-S 检验

但是, χ^2 检验的检验数无法区分在同一类别区间上有相同频数的不同分布, 类别划分对检验

表 1 出口箱集港时间与进口箱疏港时间的分组频数

Tab. 1 The frequency of export and import containers in each group

区间	进口频数	出口频数
[0,10)	2431	3034
[10,20)	1787	2665
[20,30)	1651	1988
[30,40)	1828	888
[40,50)	747	1024
[50,60)	599	787
[60,70)	362	714
[70,80)	414	616
[80,90)	481	597
[90,100]	759	1591

结果的影响也较大. 因此, 可以进一步使用 K-S 检验进行拟合优度检验. K-S 检验统计量的计算公式为:

$$D_n = \max_{1 \leq i \leq n} \{F_n(X_i) - \hat{F}(X_i)\} \quad (14)$$

式中: $F_n(X_i)$ 为经验分布函数.

(3) 后验检验

为更直观地比较各个概率分布模型的估计精度, 利用文献[17]和文献[18]提出的均方根误差(mean-square-root error, MSRE)估计概率分布模型与数据观测分布间的差异. MSRE 可表示为:

$$E_{\text{MSRE}} = \sqrt{\frac{1}{M} \sum_{i=1}^M (R_{Ei} - R_{Oi})^2} \quad (15)$$

式中: M 为区间数; R_{Ei} 和 R_{Oi} 分别为由概率分布模型得到的在第 i 个区间的理论概率和分布直方图在第 i 个区间的概率. 取区间长度为 1, 即 $M=100$.

3 集装箱码头交通需求预测模型

利用非参数核密度估计求得集疏港时间的概率密度函数后, 就能求出任意船舶任意时段到达集装箱码头的送箱集卡和提箱集卡数量. 将所有船舶的送箱集卡和提箱集卡数量叠加后, 就能得到任意时段集装箱码头的交通需求. 集装箱码头吸引交通量计算公式为:

$$D_t = \sum_{j \in V_t} D_{tj}^I + \sum_{j \in V_t} D_{tj}^E \quad (16)$$

式中: D_t 为时段 t 集装箱码头的吸引交通量; D_{tj}^I 为时段 t 集装箱码头吸引的船舶 j 的疏港交通量; D_{tj}^E 为时段 t 集装箱码头吸引的船舶 j 的集港交通量;

V_t 为时段 t 前后各一周内预计到港的船舶集合。
集装箱码头吸引的疏港交通量和集港交通量的计算公式分别为:

$$D_{ij}^I = Q_j^I \left(\hat{F}_I \left(\frac{tL - T_j^{IS}}{T_j^{IE} - T_j^{IS}} \right) - \hat{F}_I \left(\frac{(t-1)L - T_j^{IS}}{T_j^{IE} - T_j^{IS}} \right) \right) / \alpha \tag{17}$$

$$D_{ij}^E = Q_j^E \left(\hat{F}_E \left(\frac{tL - T_j^{ES}}{T_j^{EE} - T_j^{ES}} \right) - \hat{F}_E \left(\frac{(t-1)L - T_j^{ES}}{T_j^{EE} - T_j^{ES}} \right) \right) / \alpha \tag{18}$$

式中: Q_j^I 和 Q_j^E 分别表示船舶 j 由公路集疏运的进口箱和出口箱量; $\hat{F}_I(\cdot)$ 和 $\hat{F}_E(\cdot)$ 分别表示进口箱疏港时间和出口箱集港时间的累积分布函数; α 为集卡的平均载箱量; L 为时段长度。

4 算例分析

首先,分别采用基于经验法则带宽估计的核密度估计(KDE-ROT)和基于交叉验证带宽估计的核密度估计(KDE-CV)以及文献[8]和文献[9]中使用的贝塔分布、韦伯分布这两种传统概率分

布模型对出口箱集港时间和进口箱疏港时间的概率密度进行估计。进口箱疏港时间和出口箱集港时间的直方图以及各种概率分布模型的概率分布曲线如图 4 和图 5 所示,各种概率分布模型的拟合效果检验结果如表 2 所示。由图 4 和图 5 可以看出,文献[8]和文献[9]中使用的贝塔分布和韦伯分布确实能反映出出口箱集港时间和进口箱疏港时间的大致分布规律。但是,这两种传统概率分布模型均未能通过 χ^2 检验和 K-S 检验,说明对出口箱集港时间和进口箱疏港时间总体分布服从贝塔分布和韦伯分布的假定是不成立的。由图中可以看出,集疏港时间的样本分布存在多个尖峰,但是传统的概率分布模型的概率密度分布曲线均过于平滑,将很多重要的峰值掩盖了。可见,传统的概率分布模型不能准确地描述出口箱集港时间和进口箱疏港时间的分布规律。交叉验证法得到的带宽要小于经验法则法,虽然导致概率密度曲线有一定波动,但却能够更准确地反映出出口箱集港时间和进口箱疏港时间的分布规律,是四种概率分布模型中唯一能够通过 χ^2 检验和 K-S 检验的,可见带宽的选择直接关系到核密度估计能否成功;而且,KDE-CV 模型对出口箱集港时间和进口箱疏港时间的概率分布拟合效果均非常好,表明该模型具有很好的稳定性和广泛的适用性。

表 2 核密度估计模型与传统概率分布模型的拟合效果检验结果

Tab. 2 The statistical test result of non-parametric kernel density estimation and conventional parametric distribution models

航线	估计模型	χ^2	临界值	检验结果	K-S 检验	临界值	检验结果	E_{MSRE}
进口	KDE-ROT ($h_0 = 2.123\ 81$)	102.172\ 7	15.51	未通过	0.022\ 28	0.013\ 39	未通过	0.004\ 28
	KDE-CV ($h_v = 0.002\ 25$)	0.001\ 23	15.51	通过	0.000\ 41	0.013\ 39	通过	0.000\ 04
	贝塔分布	1\ 661.693\ 35	14.07	未通过	0.130\ 59	0.013\ 39	未通过	0.006\ 45
	韦伯分布	2\ 272.900\ 41	14.07	未通过	0.045\ 19	0.013\ 39	未通过	0.005\ 73
出口	KDE-ROT ($h_0 = 2.128\ 4$)	98.421\ 55	15.51	未通过	0.022\ 5	0.011\ 53	未通过	0.002\ 89
	KDE-CV ($h_v = 0.083\ 13$)	0.048\ 36	15.51	通过	0.009\ 01	0.011\ 53	通过	0.000\ 21
	贝塔分布	1\ 693.784\ 9	14.07	未通过	0.120\ 72	0.011\ 53	未通过	0.006\ 06
	韦伯分布	6\ 177.352\ 49	14.07	未通过	0.059\ 04	0.011\ 53	未通过	0.004\ 87

现将 1 周划分为 84 个时段,利用式(16)~(18)分别计算 KDE-CV 模型、贝塔分布、韦伯分布 3 种概率分布拟合方法下,各个时段集装箱码

头的吸引交通量,计算结果如图 6 所示。可以看出,KDE-CV 模型的预测精度明显高于贝塔分布、韦伯分布这两种传统参数估计模型,除了个别实

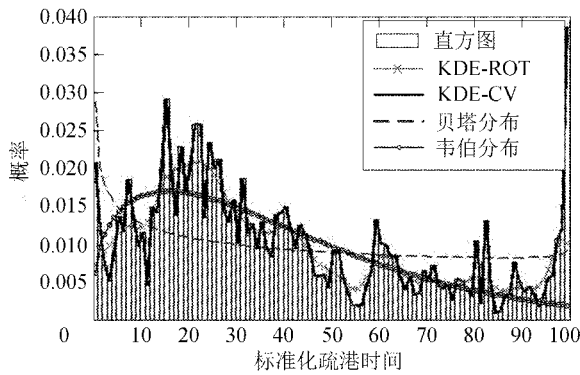


图 4 进口箱疏港时间的直方图和概率密度曲线

Fig. 4 The histogram and probabilistic density curve of import containers' pick up time

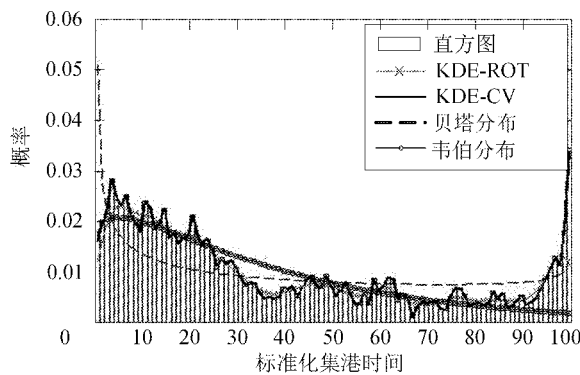


图 5 出口箱集港时间的直方图和概率密度曲线

Fig. 5 The histogram and probabilistic density curve of export containers' delivery time

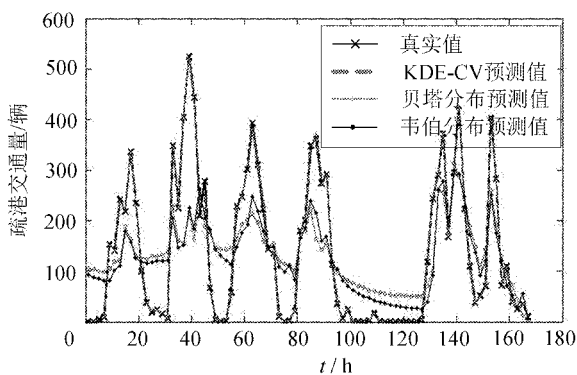


图 6 吸引交通量预测数据与实际数据比较

Fig. 6 Comparisons between forecast data and actual data of attracted traffic volume

际的吸引交通量很少的时段外,误差均在 5% 以内,平均误差为 2.34%;而传统参数估计模型预测结果的误差均在 10% 以上。可见,KDE-CV 模型能够提供更准确的吸引交通量预测。以此为基础,不仅能够更加准确地预测集疏港通道的交通量,

而且能够根据预测的作业任务量(交通量)更好地配置和调度码头资源。

5 结 论

本文基于非参数核密度估计理论建立集装箱集疏港时间的概率分布模型以及集装箱码头交通需求预测模型,并利用该模型对 DCT 码头连续 2 周时间内 94 艘次集装箱船的 13 906 个出口箱的集港时间和 10 322 个进口箱的疏港时间的概率分布进行研究。结果表明:(1)非参数核密度估计的拟合效果明显优于传统的参数模型估计方法,可以通过拟合优度检验;(2)非参数核密度估计模型对出口箱集港时间和进口箱疏港时间的概率分布均有很高的拟合精度,表明该模型具有很好的稳定性和广泛的适用性;(3)本文建立的基于非参数核密度估计的集装箱码头交通需求预测模型具有较传统参数模型更高的预测精度。本文提出的方法保证了最终得到的集装箱集疏港时间的概率分布规律的准确性、适应性和高效性,可为集装箱码头基础设施规划、集疏港通道的道路交通管理、码头资源优化配置和调度等问题的研究提供更准确的交通量和作业任务量预测。

参考文献(References):

- [1] KOUROUNIOTI I, POLYDOROPOULOU A. Identification of container dwell time determinants using aggregate data[J]. International Journal of Transport Economics, 2016, 44(4): 567-588.
- [2] KOUROUNIOTI I, POLYDOROPOULOU A, TSIKLIDIS C. Development of models predicting dwell time of import containers in port container terminals - an artificial neural networks application [J]. Transportation Research Procedia, 2016, 14:243-252.
- [3] 杨龙海,章锡俏,李社教,等.深圳盐田港保税区货运交通需求预测[J].城市交通,2007,5(2):51-54. YANG H L, ZHANG X Q, LI S J, et al. The freight traffic demand forecasting of Shenzhen Yantian Harbor Free Trade Zone [J]. Urban Transport of China, 2007, 5(2):51-54. (in Chinese)
- [4] BOER C A, VERBRAECK A, DE A, et al. Distributed E-services for road container transport simulation [C]// Proceedings 15th European Simulation Symposium. 2003:541-550.
- [5] 滕藤,曾庆成,陈文浩.集装箱码头集卡到达时间预

- 测模型[J]. 武汉理工大学学报(交通科学与工程版), 2015, 39(6):1221-1224.
- TENG T, ZENG Q C, CHEN W H. Prediction model for truck arrival time in container terminals[J]. Journal of Wuhan University of Technology (Transportation Science & Engineering), 2015, 39(6):1221-1224. (in Chinese)
- [6] 张莎莎, 庞明宝, 杜宏亮, 等. 基于船期表的一种集装箱码头闸口交通需求智能预判方法[J]. 交通科技, 2014 (5):148-151.
- ZHANG S S, PANG M B, DU H L, et al. An intelligent traffic demand forecasting method for container terminal based on line schedule [J]. Transportation Science & Technology, 2014 (5):148-151. (in Chinese)
- [7] 陈家源, 郑仁环. 港口通过能力系统理论研究——港口交通流的分布与生产状态的分析方法[J]. 大连海事大学学报, 1993, 19(3):297-309.
- CHEN J Y, ZHENG R H. System theories research of the port's throughput capacity [J]. Journal of Dalian Maritime University, 1993, 19(3):297-309. (in Chinese)
- [8] 杨忠振, 陈刚, 陈康. 基于码头作业形态的港区道路交通需求预测[J]. 大连海事大学学报, 2009, 35(4):35-38.
- YANG Z Z, CHEN G, CHEN K. Traffic demand forecasting model for large port based on terminal operation situation [J]. Journal of Dalian Maritime University, 2009, 35(4):35-38. (in Chinese)
- [9] YANG Z, CHEN G, MOODIE D R. Modeling road traffic demand of container consolidation in a Chinese port terminal [J]. Journal of Transportation Engineering, 2010, 136(10):881-886.
- [10] 郭子坚, 方泽兴, 王文渊, 等. 集装箱港区集疏货运交通随机时变特性仿真研究[J]. 港工技术, 2017, 54(5):11-15.
- GUO Z J, FANG Z X, WANG W Y, et al. Simulation research on freightage changing randomly with time in container harbor [J]. Port Engineering Technology, 2017, 54(5):11-15. (in Chinese)
- [11] 薛留根. 应用非参数统计[M]. 北京:科学出版社, 2013.
- [12] 孙志华. 非参数与半参数统计[M]. 北京:清华大学出版社, 2016.
- [13] PARZEN E. On estimation of a probability density function and mode [J]. The Annals of Mathematical Statistics, 1962, 33(3):1065-1076.
- [14] 徐玉琴, 张扬, 戴志辉. 基于非参数核密度估计和Copula函数的配电网供电可靠性预测[J]. 华北电力大学学报, 2017, 44(6):14-19.
- XU Y Q, ZHANG Y, DAI Z H. Power supply reliability prediction of distribution network based on nonparametric kernel density estimation and copula function [J]. Journal of North China Electric Power University, 2017, 44(6):14-19. (in Chinese)
- [15] RUDEMO M. Empirical choice of histograms and kernel density estimators [J]. Scandinavian Journal of Statistics, 1982, 9(2):65-78.
- [16] BOWMAN A W. An alternative method of cross-validation for the smoothing of density estimates [J]. Biometrika, 1984, 71(2):353-360.
- [17] QIN Z, LI W, XIONG X. Estimating wind speed probability distribution using kernel density method [J]. Electric Power Systems Research, 2011, 81(12):2139-2146.
- [18] 颜伟, 任洲洋, 赵霞, 等. 光伏电源输出功率的非参数核密度估计模型[J]. 电力系统自动化, 2013, 37(10):35-40.
- YAN W, REN Z Y, ZHAO X, et al. Probabilistic photovoltaic power modeling based on nonparametric kernel density estimation [J]. Automation of Electric Power Systems, 2013, 37(10):35-40. (in Chinese)