

基于随机森林的码头集装箱堆存时间的预测

胡水源¹ 陈以¹ 劳钰莎² 蒋志³

1. 桂林电子科技大学; 2. 广州中医药大学; 3. 南宁华资科技有限公司 广西桂林 541004

摘要: 码头集装箱的堆存时间是码头作业效率的一个重要指标,也是码头规划集装箱作业决策的主要参考。这项研究提出了码头集装箱卸船时箱号、箱状态、目的港和载货港等多个指标作为参数输入,通过决策树、随机森林等多个机器学习模型来预测到达码头的集装箱的堆存时间。这项研究结果表明,随机森林预测模型相比其他模型具有较好的泛化能力,并且发现目的港口和航线对集装箱的堆存时间的预测结果具有较大的影响。

关键词: 机器学习; 随机森林; 堆存时间; 短期预测

1. 引言

随着国内外贸易的发展和一带一路的推行,在进口集装箱到达码头后,集装箱进入堆场的随意性和无序性,在后续的作业过程中,针对闸口收发和装卸船等作业时,会对堆场的集装箱进行翻箱操作,会导致可能堆存时间较短的集装箱被翻箱至下层,从而造成后续的成本的增加和装车效率的降低。若能较好的预测卸船时集装箱的堆存时间,对堆存时间长的集装箱做单独摆放,堆存时间短的放置上层,那么就可以从侧面降低码头的翻箱率,有助于减少码头资源的浪费和装车时间,提高装车效率,促进码头经济发展。

2. 码头集装箱堆存时间的预测现状

目前,国内外学者对于时间预测的研究主要集中在生产和运输工具上,而对于堆存时间预测的预测相对较少。在时间预测方面,传统的预测时间方法^[1]有随机森林预测法、支持向量机方法、BP神经网络方法、卡尔曼滤波方法、极限学习机方法、K最近邻算法、梯度回归树方法等。而在码头时间预测方面,赵宇迪^[1]提出了考虑AIS信息的船舶到港时间预测模型的研究,结合静态数据、动态数据来预测船舶到港时间,来进行堆场后续规划。刘井田^[2]提出了集疏运平台中集卡到达时间预测方法的研究,建立了基于K近邻回归的预测模型,通过对集装箱从其它场站或码头到达本码头的行程时间进行预测,方便集装箱到达码头前做好规划。

3. 实验结果分析

在这次研究中,以广州某港口码头为例,收集了2020年1月1日-2021年8月25日通过该码头卸船数据总计945687条卸船数据,并根据码头要求可控误差为正负一天进行预测。通过前期的数据分析,发现码头集装箱的箱号,持箱人,空重箱,箱型,载货港,目的港,进口航线,进口航次,货物类型,超码箱标识和货物状态

等指标均对堆存时间有影响,其中目的港口和进口航线影响较大。考虑时间对其数据的影响,因此采取随机选取数据,前80%作为堆存时间的训练集,后20%作为堆存时间的测试集。

在该数据集中,本研究采用了不同的机器学习模型进行预测,分别使用了Adaboost回归,Tree回归,Bagging回归,ExtraTree回归,随机森林回归,并进行了实验对比。在实验中,总共进行了五组比较实验,最终结果如表1所示。

表1 预测模型实验结果

	Adaboost 回归	Tree 回归	Bagging 回归	ExtraTree 回归	随机森林 回归
准确率	15.14%	83.99%	83.89%	82.39%	84.31%

决策树^[1]是一种用来表示判定规则的树结构。通过对训练样本集进行训练,构造出一棵可以表示一定规则的决策树,该树对样本空间进行了划分。当使用决策树对未知样本预测时,该算法利用已生成的树,从根结点开始对样本的属性测试其值,并顺着分枝向下移动,直至达到某个叶结点为止。此叶结点代表的类即为该样本的分类结果。

Bagging是最常用和最实用的集成学习算法之一,旨在结合许多弱分类器获得强分类器,通过建立许多独立的训练数据集生成一些独立的分类器,然后将它们组合起来,以多数投票构建最终的分类器。

ExtraTree回归使用的是所有的样本,算法和随机森林算法类似,但是选择的特征是随机的,因为分裂是随机的,

随机森林该方法结合Bagging和随机选择的特征被独立引入,以构建具有可控变化的决策树集合。

另外,将测试集随机选取50个箱号进行预测,随机森林回归预测结果如图1所示,其中蓝点表示真实值,

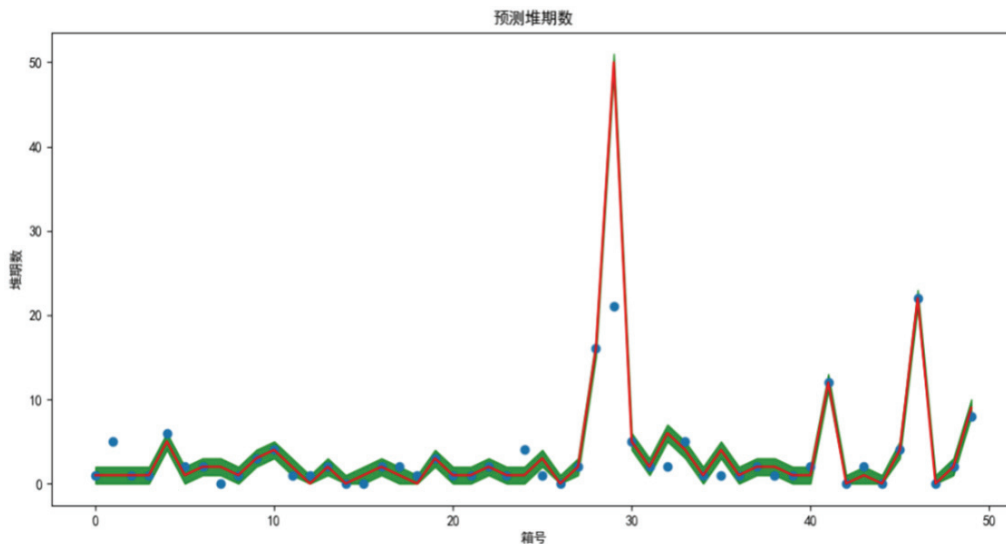


图1 随机森林回归预测结果

红色线拐点为预测值，绿色区间表示误差为正负1，及蓝点在绿色区间则表示预测结果正确，反之预测失败。

4. 实验结论

随着码头进口集装箱的数量增加，码头堆场的空间，导致了码头集装箱翻箱率高，提箱效率低，这将造成码头经济的损失和成本的增加。因此，根据进口集装箱的基本信息，准确预测未来集装箱的堆存时间将有助于减少码头翻箱作业和装卸时间，合理决策货物存放位置和提箱预约请求。在这项研究中，基于随机森林模型的方法建立了一个集装箱堆存时间的预测模型，以预测码头进口集装箱的堆存时间。此次研究总结如下：

(1) 本研究在模型中使用随机森林预测模型来获取更准确的码头进口集装箱堆存时间的预测。并与其他预测模型相比较，随机森林预测的准确率为84.31%。因此，随机森林相比其他预测模型在码头进口集装箱堆存时间的预测中具有较高的准确性与应用价值。

(2) 虽然随机森林预测模型准确率高于本研究中其它的预测模型，在预测方面具有令人满意的性能，但是仍有一些地方需要改进。例如，缩短预测的误差，将误差缩短为正负半天。考虑天气对堆期的影响，例如：在某航线中出现台风等特殊情况，也会对结果产生干扰。

参考文献：

- [1] 赵宇迪. 考虑 AIS 信息的船舶到港时间预测模型研究[D]. 大连理工大学, 2018.
- [2] 滕藤. 集装箱码头外部集卡到港量预测模型[D]. 大连海事大学, 2017.
- [3] 刘炳春, 张鹏. 基于机器学习的港口集装箱吞吐量预测[J]. 中国储运, 2021 (03): 123-124.
- [4] 张弛, 王萍, 苏佳山, 程冬梅. 基于机器学习算法的干眼预测模型研究[J]. 国际眼科杂志, 2021, 21 (09): 1644-1648.