

大数据环境下船舶交通客流量实时预测模型

李俊翰

(重庆电子工程职业学院人工智能与大数据学院, 重庆 401331)

摘要: 针对当前船舶交通客流量实时预测方法受到冗余数据的影响, 导致预测结果精度较差, 因此提出大数据环境下船舶交通客流量实时预测模型。从进、出站客流均值和极值两方面, 分析船舶交通客流量时间序列特征, 引入决策表简约信息属性, 剔除冗余数据, 构建船舶交通客流量实时预测模型, 分析正常和异常船舶交通客流量转换方程, 了解短期内客流量变化情况, 简化实时预测模型, 确定时间序列中各个元素权重系数, 实现船舶交通客流量实时预测。实验结果表明, 该模型预测的客流量与实际客流量一致, 预测结果精度较好。

关键词: 大数据环境; 船舶交通; 客流量; 实时预测模型

中图分类号: TP399 文献标识码: A

文章编号: 1672-7649(2021)2A-0025-03 doi: 10.3404/j.issn.1672-7649.2021.2A.009

Real-time forecasting model of ship traffic and passenger flow under big data environment

LI Jun-han

(School of AI and Big Data, Chongqing College of Electronic Engineering, Chongqing 401331, China)

Abstract: In view of the problem that the current real-time prediction method of ship traffic passenger flow is affected by redundant data, which leads to the poor accuracy of prediction results, a real-time prediction model of ship traffic passenger flow in big data environment is proposed. This paper analyzes the characteristics of time series of ship traffic passenger flow from the mean value and extreme value of inbound and outbound passenger flow, introduces the simplified information attribute of decision table, eliminates redundant data, constructs the real-time prediction model of ship traffic passenger flow, analyzes the conversion equation of normal and abnormal ship traffic passenger flow, understands the change of passenger flow in short term, simplifies the real-time prediction model, and determines the time series the weight coefficient of each element is used to realize the real-time prediction of ship traffic passenger flow. The experimental results show that the passenger flow predicted by the model is consistent with the actual passenger flow, and the prediction result has good accuracy.

Key words: big data environment; vessel traffic; traffic; real-time prediction model

0 引言

船运客流预测研究是船运安全的基础, 是船运工程规划的数据支撑^[1]。通过建立预测模型, 利用回归分析方法对船舶客流进行预测, 结合相关系数分析二者的相关性, 确定其线性关系; 将神经网络方法应用于预测中, 建立了多层前向神经网络预测模型, 并结合误差修正算法, 得到了精确预测结果^[2]。上述2种回归分析和神经网络预测方法对不同船舶新线路开通、运营中心迁移的交通变化反应较慢, 预测结果精度较差。因此, 建立了在大数据环境下的船舶交通流量实时预测模型, 以帮助相关船舶交通部门及时调整船舶

调度信息。

1 大数据环境下船舶交通客流量时间序列特征分析

大数据环境下, 交通客流时间序列特征对整个客流的影响主要体现在2个方面:

1) 进出站客流均值: 即每日进出站客流从运营开始到结束的平均时间^[3]。对其价值的统计、分析和预测, 反映了研究站点的吸引力和整个船舶交通网络的运行水平。

2) 进出站客流最大值: 从营运开始到营运结束的

收稿日期: 2020-10-19

基金项目: 教育部2018年度人文社科规划基金项目(18YJA880061)

作者简介: 李俊翰(1984-), 男, 硕士, 讲师, 研究方向为软件技术及数据分析与应用。

入境客流平均最高值和最低值，船舶交通进出区最大客流是其运营管理的重要指标之一。站内日平均最大客流与站外主客源运行状况有显著相关性。

2 信息属性简约

信息和决策表是粗糙集的核心组成部分，利用它们之间决策属性关系，可以有效地去除冗余数据^[4]。

定义冗余属性，假设 $S = \langle U, A, C, f \rangle$ 是一个信息系统，令 $a \in P$, $INDP = IND(P - \{a\})$ 。则可以判定在当前信息集合中， a 的属性在 P 中多余。这是个不必要的属性，需要相对减少。相对论的简化 $S = \langle U, A, C, f \rangle$ 是一个决策表， C 中所有约减属性 D 记做是 $Red(D)$ 且 $Red(D) = \cap Red(D)$ 。

结合当前数据粗糙集，确定决策表，该表集合为： $S = (U, C \cup D, V, f)$ ，根据表中 $\forall B \subseteq C, \forall \beta \in C, \forall a \in C - B$ ，可以得出以下公式：

$$\text{sig}(a, B; D) = \gamma_{IND(B \cup \{a\}}(D) - \gamma_{IND(B)}(D) = \frac{POS_{B \cup \{a\}}(D)}{U} \quad (1)$$

式中： a 表示数据条件； B 表示决策程度； D 表示船舶交通客流量。

$$\text{sig}(\beta, C; D) = \gamma_c(D) - \gamma_{c-\beta}(D), \quad (2)$$

式中， β 表示数据条件； C 表示相对决策程度。

可以看出，当前数据属性的重要性在粗糙集决策表属性约简过程中起着重要作用。确定有效的重要功能，然后根据实际问题需求进行缩减。

3 客流量实时预测模型构建

以约简信息属性为支撑，利用状态转移方程和观测方程来估计离散控制系统的真实状态，利用连续测量信息来修正预测值，以减少期望输出和实际输出之间的误差。使用如图 1 所示的大数据预测方法建立客流预测模型。

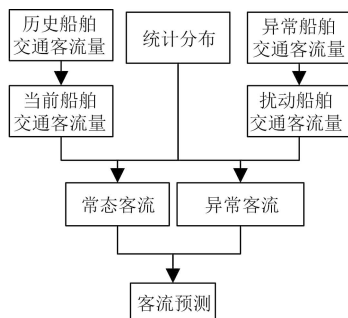


图 1 客流预测模型
Fig. 1 Passenger flow prediction model

在图 1 的预测模型中，缺少历史客流、船舶交通客流的时序特征、环境变化数据、常规客流、干扰客流、正常客流和异常客流是无法进行预测的。船舶交通客流量统计综合考虑了上述因素，在此基础上，建立了正常、异常船舶客流空间状态转换方程，并对正常、突发船舶客流进行短期预测。

设在 n 个 $h_i(t)$ 时间序列内存在正常客流，客流量实时预测模型可表示 $H(t)$ 在 $h(t)$ 上映射，公式为：

$$H(t) = \psi(h(t)) = \psi(h_1(t), h_2(t), \dots, h_n(t)), \quad (3)$$

式中， $\xi(\cdot)$ 为映射关系符号。由于 $h_i(t)$ 数据量大，存在冗余数据，所以通过信息简约处理后，可得到 $h'_i(t)$ ，由此得到的简化实时预测模型为：

$$H(t) = \sum_{i=1}^n \lambda_i(t) \psi(h'_i(t)), \quad (4)$$

式中， $\lambda_i(t)$ 表示时间序列中各个元素权重系数。

船舶交通客流的实时预测过程见图 2。

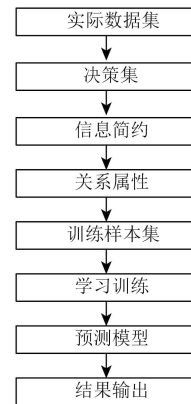


图 2 模型预测实现流程

Fig. 2 Model prediction implementation flow

可知，在简约信息支持下，不断深入分析船舶交通大数据。对于运营经理而言，分析相关有价值的信息可以优化并指导生产操作；而对于乘客而言，则可以利用交通信息系统提供的交通服务信息，方便其出行。并且以此平台为用户提供新的信息需求，实现对系统结构、业务模块和功能域的不断完善和更新，为乘客提供更加实时准确的信息。

4 实验分析

为了验证大数据环境下船舶交通客流量实时预测模型构建合理性与实际应用可行性，以长江苏通大桥客流量进行实验验证分析。

4.1 实验数据设置

统计长江苏通大桥实际客流量，如表 1 所示。

表 1 长江苏通大桥实际客流量/人
Tab. 1 Actual passenger flow/person of sutong bridge of the yangtze river

时间/h	2018/10/01	2018/10/02	2018/10/03
2	50	56	60
4	60	58	62
6	70	75	65
8	70	75	69
10	80	83	70
12	85	85	78
14	85	89	78
16	90	92	83
18	92	95	85
20	95	100	90
22	100	100	95
24	100	100	99

可知，在 2018/10/01，实际客流量为 100 人，达到客流量上限；在 2018/10/02，实际客流量为 100 人，达到客流量上限；在 2018/10/03，实际客流量为 99 人。

4.2 实验结果与分析

分别使用回归分析法、神经网络法和大数据环境下预测模型对长江苏通大桥实际客流量预测，预测结果如图 3 所示。

可知，使用回归分析法在 2018/10/01 ~ 2018/10/03 期间，2018/10/01 客流量与实际客流量相差较大，其他 2 个时间也存在一定差量；使用神经网络法在 2018/10/01 ~ 2018/10/03 期间，也是在 10/01 客流量与实际客流量相差较大；而使用大数据环境下预测模型与实际客流量人数一致。由此说明，使用该方法预测结果更加精准。

5 结 语

基于粗糙集预测机制，设计了大数据环境下船舶交通客流量实时预测模型。此模型能集中筛选当前顾客信息的有效属性，并以此为训练样本重新组织，产生预测结果。实验结果表明，该模型预测流量接近实际流量，预测结果精度较好。因此，所建模型具有较

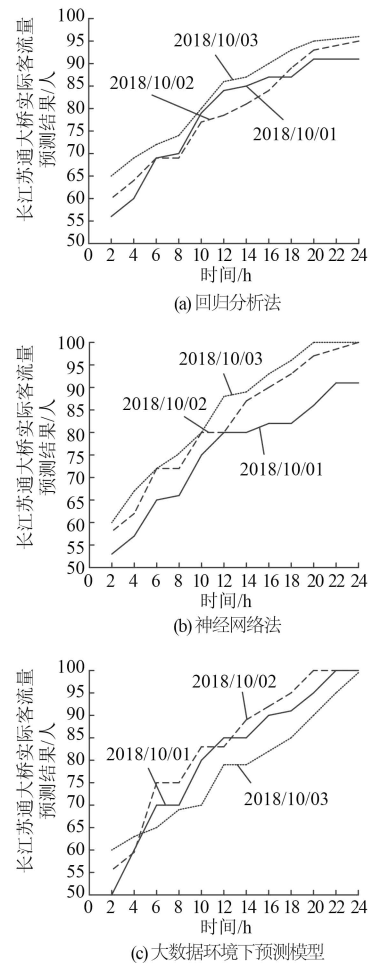


图 3 3 种方法长江苏通大桥客流量预测

Fig. 3 Passenger flow prediction of sutong bridge in the yangtze river by three methods

高的实时性，对船舶流量的预测具有一定的可行性。

参考文献：

- [1] 王雪菲, 丁维龙. 面向高速公路大数据的短时流量预测方法 [J]. 计算机应用, 2019, 39(1): 93-98.
- [2] 刘艳丽, 赵卓峰, 丁维龙, 等. 基于高速收费大数据的短时交通流量预测方法 [J]. 计算机与数字工程, 2019, 47(5): 155-160+179.
- [3] 蔡晓禹, 谭宇婷, 雷财林, 等. 交通大数据环境下短时交通流量预测研究 [J]. 铁道运输与经济, 2018, 40(8): 88-93.
- [4] 杨帆, 何正伟, 何帆. 基于深度网络和船舶交通流的航道水深预测方法研究 [J]. 武汉理工大学学报: 交通科学与工程版, 2019, 43(1): 130-135.