

文章编号: 1000-4653(2010)02-0060-03

数据挖掘在海上交通特征分析中的应用研究

潘家财^{1,2}, 邵哲平², 姜青山¹

(1. 厦门大学 信息科学与技术学院, 福建 厦门 361011; 2. 集美大学 航海学院, 福建 厦门 361021)

摘要: 数据挖掘(Data Mining)就是从大量数据中获取有效的、新颖的、潜在有用的、最终可理解的模式的过程。而海量的船载 AIS 信息集合蕴藏着大量的海上交通特征,利用数据挖掘理论与技术,以厦门港主航道的大型船舶 AIS 信息为对象,从空间、时间、概率分布三个方面,对 AIS 信息集合蕴藏的海上交通特征中反映船舶航向和航速变化率进行挖掘研究,获得了厦门湾内主航道通航环境空间和时间特性以及航向航速变化率概率分布特点。该研究方法为海量 AIS 信息研究提供新的思路,利用所挖掘的知识可以分析水域的通航环境,为港口主管部门的通航环境管理等提供理论依据。

关键词: 水路运输; 数据挖掘; 海上交通; 船舶 AIS 信息; 通航环境

中图分类号: U 675.7; TP391 文献标志码: A

Application of Data Mining Technology in Analysis of Marine Traffic Characteristics

PAN Jia-ai^{1,2}, SHAO Zhe-ping², JIANG Qing-shan¹

(1. School of Information Science and Technoledge, Xiamen University, Xiamen 361021, China;
2. College Navigation, Jimei University, Xiamen, 361005, China)

Abstract: Data mining is a process of analyzing huge data from different perspectives and extracting useful information. The technology and theory of Data Mining play an important role in find out the characteristics of marine traffic through analyzing huge AIS data broadcast by ships and collected into database. A novel method for the analysis of marine traffic situation is developed based on the theory and technology of data mining. With this analysis method ship course change rates and speed change rates are mined in three respects (the space, time and distribution) and the analysis of marine traffic situations in Xiamen main fairway is carried out. It is proved that this method is practical and useful. Knowledge mined by this analysis method can give advices for marine traffic decision-making and management of marine authorities.

Key words: waterway transportation; data mining; marine traffic; shipborne AIS data; marine traffic situation

海上交通^[1]是指定区域内船舶运动的组合与船舶行为的总体。通过分析海上交通特征的规律,对相关机关制定管理措施、发展规划,航道设计,航标设置,定线制设计等都具有指导意义。

从 2002 年至今,大量船舶均安装有船载 AIS (Automatic Identification System) 设备,它对船舶安全航行和防止海洋污染起到了重要的作用。AIS 信息包含丰富的船舶动态和静态信息,通过 AIS 信息采集技术,可获得大范围海域的大量的船舶同时

连续不断播发 AIS 信息,这些信息构筑了数据量巨大信息源,其中蕴藏着大量的海上交通特征^[2-6]。

数据挖掘(Data Mining)就是从大量数据中获取有效的、新颖的、潜在有用的、最终可理解的模式的过程。

简单的说,数据挖掘就是从大量数据中提取或“挖掘”知识。它是一门涉及面很广的交叉学科,包括机器学习、数理统计、神经网络、数据库、模式识别、粗糙集、模糊数学等相关技术。数据挖掘的主要

收稿日期: 2010-03-31

基金项目: 福建省自然科学基金计划资助项目 (A0640007); 福建省教育厅科技项目 (JA07134); 厦门市科技局重点科技项目 (3502Z20083023)

作者简介: 潘家财(1975-),男,福建南安人,博士研究生,从事数据挖掘、交通信息工程及控制、航海技术研究。E-mail: panjiacai@163.com.

© 1994-2012 China Academic Journal Electronic Publishing House. All rights reserved. <http://www.cnki.net>

步骤有:数据清洗、数据集成、数据转换、数据挖掘、模式评估、知识表示^[7-8]。

本文采用数据挖掘理论和方法,对大量收集到的船舶 AIS 数据进行整理、转换和清理,提取反映海上交通特征的数据内容,为深入研究海上交通,为港口主管部门的通航环境管理等提供理论依据。

1 表征海上交通状况的基本特征

反映海上交通的主要特征包含有船舶的交通量、船舶密度分布、船舶航迹分布和船舶航向和航速变化率等。船舶的交通量是反映水域船舶交通规模和繁忙程度。可按船舶用途和船舶种类,按国籍、船舶大小、航区、航线等方法进行分类。船舶密度分布是水域船舶的空间分布。它可具体地反映船舶的空间分布的实况。船舶航迹分布是船舶运动路线的空间分布。它显示海上交通路线的特征或交通形式。船舶航向和航速变化率表征船舶的运动变化的剧烈程度。船舶在水上交通环境良好、安全的水域航行,一般情况是航向和航速平稳。而在避让船舶时,主要采取转向和/或变速措施。如果水域的交通状况恶劣,则船舶需要不断改变航向和/或航速来避让其它船舶或危险物。因此,航行中船舶的航向和航速改变幅度大小、频繁程度则直接地反映了该水域的通航环境状况。特别是船舶航向改变幅度大且频繁,则反映出了该水域可能存在较大的安全隐患。因此,本文选取航向和航速的单位时间变化率作为主要研究对象。

2 数据挖掘技术的分析步骤

2.1 研究对象的选取和清理

海量的船舶 AIS 数据是存储在数据库中,在数据挖掘中必须根据研究对象来选取和清理相关信息。例如可根据厦门港的情况选取研究的地理位置

范围为纬度 $24^{\circ} 19' . 26N$ 至 $24^{\circ} 26' . 04N$, 经度 $118^{\circ} 02' . 64E$ 至 $118^{\circ} 13' . 20E$ 所构成的水域,即厦门湾主航道从九节礁附近到鼓浪屿南边航道分叉的 10.3 n mile 航段进行研究。研究的水域范围及相关的船舶航迹分布如图 1 所示。

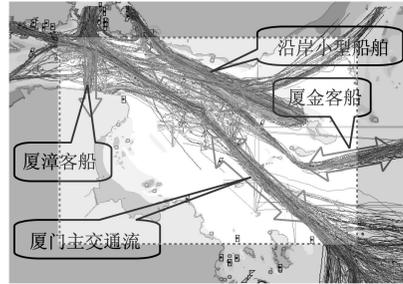


图 1 厦门湾水域船舶航迹分布示意图

Fig. 1 Research Area and Ship track distribution plotting

从图 1 中可知,厦门湾内的主要交通流有厦门主交通流(大型船舶)、厦金客船(小型客船)、沿岸小型船舶、厦漳客船。

2.2 数据处理

由于数据库中的船舶航向和速度是采集时间点的实时值,因而需要变换,以反映航向和航速的改变情况。采用式(1)获取航向和航速的变化率。

$$\text{航向或航速变化率} = \frac{\text{相邻时间航向或航速差值}}{\text{相邻时间间隔}} \quad (1)$$

为便于观测和运算,获得航向和航速的平均变化率的单位分别为度每分钟($^{\circ} \cdot \text{min}^{-1}$)和节每分钟($\text{kn} \cdot \text{min}^{-1}$)。

数据选取的时间范围为 2008 年 1 月、3 月、5 月、7 月、9 月和 11 月共 6 个月,每个月选中其中的 1—5 日共 5 d 的船舶 AIS 动态数据和整个 2 月份数据。选择的船舶大小为船长在 250~ 340 m 之间的大型船舶。根据式(1)从数据库中可得出符合上述条件的船舶 AIS 数据统计情况如表 1 所示。

表 1 AIS 数据统计情况

Tab. 1 Details of AIS data Sample

月份	1	3	5	7	9	11	合计	2月整月
AIS 数据记录数/艘	1 172	1 049	1 291	876	966	693	6 047	6 0454
统计出的船舶数量/艘	12	6	6	10	10	6	50	172

3 航向和航速变化的空间分布

根据上述的约束条件和计算公式,将从船舶 AIS 数据库中计算出航向和航速变化的空间分布结果标示在电子海图中。其航速和航向变化率情况如

图 2~ 图 3 所示。

从图 2 中可见,在 A 点、B 点之间水域、C 点附近水域航速变化较大,比较有普遍性,出现较多的异常值。在 A 点~ B 点之间水域,通常大型船舶进入海沧港区前减速,且在主航道航行的大型船舶交通

流和厦漳客船的交通流在此交汇,大型船舶需要较多的变速来避让;且B点则是厦门港引航员上下船地点,因而B点周边船舶的变速也较为频繁。

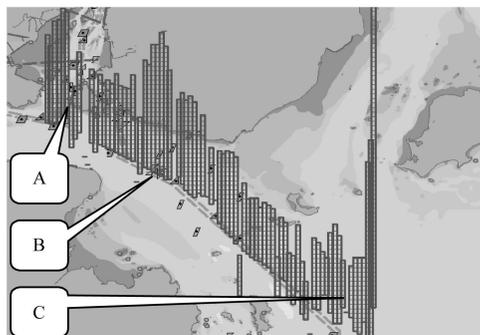


图2 航速变化率空间分布

Fig. 2 Spatial Distribution of Change Rate of Speed

而往来厦门和平码头和漳州的数量众多的小型船舶穿越A点~B点之间水域,影响大型船舶通航,为保证安全,大型船舶需要降速通行。而在C点,则是由于航道两侧有小岛屿,航道宽度变窄,船舶交通相对拥挤,所以船舶需要降速避让过往船舶。

从图3中可见,在A点、B点和C点附近,航向变化较为普遍。分析如下:

在A点附近,由于大型船舶为避让往来厦门和平码头和漳州的小型船舶而频繁改变航向,并且大型船舶在此处进行改向前往海沧或东渡港区;在B点航向变化大,则是该处有厦门的3#锚地,大型船舶本身降速改向到锚地抛锚或避让进出锚地其他船舶。C点有一个较大的航道转向点,在C点附近水域,则是由于在九节礁附近航道较窄,各方向来的船舶到此都要调整航向,以便进出厦门湾主航道。

综上所述,厦门湾主航道上水上交通环境较为复杂的区域有A和B点之间水域和C点附近水域。这些水域需要主管机关加强管理,而过往船舶则应加强了望,谨慎航行,以确保安全。



图3 航向变化率空间分布

Fig. 3 Spatial Distribution of Change Rate of Course

4 航向航速变化率分布规律分析

为简化计算,以航向变化率($1^\circ \cdot \text{min}^{-1}$)为间隔把本文的研究数据集分为16个小数据集,并统计各个小数据集内数量,获得各个航向变化率区间内的统计数量,即航向变化率的频率。而航速变化率则是以 $0.12 \text{ kn} \cdot \text{min}^{-1}$ 为间隔进行统计分析。航速变化率频率统计分布如图4所示,航向变化率频率统计分布如图5所示。从这些统计图清晰可见,航向和航速的变化率频率均服从指数分布。

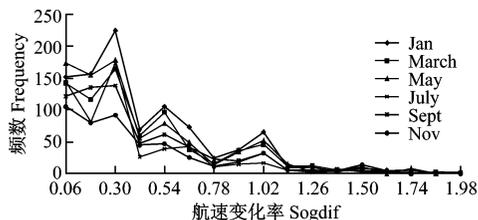


图4 航速变化率频率统计分布图

Fig. 4 Probability Distribution of change rate of speed

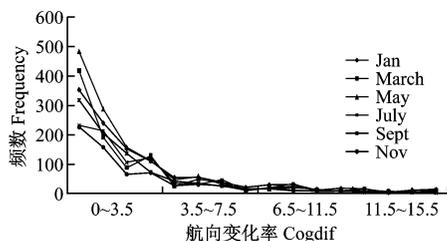


图5 航向变化率频率统计分布图

Fig. 5 Probability Distribution of change rate of course

为进一步对航向航速变化率的分布特性进行分析,利用SAS软件^[9-10]对上节的航向和航速变化率频率统计量之百分率进行指数拟合,得出航速变化率频率分布概率如式(2)所示,航向变化率频率分布概率如式(3)所示。

SOGDiff_Distr=

$$0.2529 * \text{EXP}(-1.9857 * \text{sogdiff}) \quad (2)$$

式(2)中,SOGDiff_Distr是航速变化率频率分布概率,sogdif是航速变化率频率。

COGDif_Distr=

$$0.3645 * \text{EXP}(-1.05429 * \text{cogdiff}) \quad (3)$$

式(3)中,COGDif_Distr是航向变化率频率分布概率,cogdif是航向变化率频率。

5 结语

由于AIS信息量巨大,特别是大范围的AIS信息数据库,如果没有适当的挖掘理论和方法,对理解和利用AIS信息就会存在极大的困难。本文阐述

(下转第73页)

- 略[M].北京:人民交通出版社,2006.
- [4] 王嘉波,崔志超.浅析智能运输系统及其发展方向[J].交通科技与经济,2006(1):35-38.
WANG Jia-Bo, CUI Zhi-Chao. Discussing the intelligent transport system and its development trend[J]. Communications Science and Economic. 2006(1):35-38.
- [5] 王笑京.智能运输系统(ITS)在中国的发展与研究[J].中国公共安全,2005(8):88-89.
WANG Xiao-Jing. The development of ITS in China [J]. China Public Security, 2005, 8: 88-89.
- [6] 张宁,高更君,林慧丹.基于ITS技术的现代物流体系的构建[J].物流技术,2004(3):30-34.
ZHANG Ning, GAO Geng-jun, LIN Hui-Dan. Construct Modern Logistics Architecture Based on ITS Technique[J]. Logistics Technology. 2004(3):30-34.
- [7] 胡明伟,史其信.智能运输系统(ITS)评价方法研究[J].公路交通科技,2001(10):24-26.
HU Ming-Wei, SHI Qi-Xin. Study of Intelligent Transport Systems Evaluation Methods[J]. Journal Of Highway And Transportation Research And Development. 2001(10):24-26.
- [8] 冀翔,张金财.智能运输系统——未来交通运输的发展方向[J].黑龙江交通科技,2007(9):38-40.
JI Xiang, ZHANG Jin-Cai. ITS——The Future Trend Of Transportation And Communication[J]. Communications Science And Technology HeiLongJiang, 2007(9):38-40.
- [9] 杨敏,李文勇,徐建闽.基于DHGF算法的智能运输系统(ITS)评价方法[J].桂林电子科技大学学报.2007(3):61-64.
YANG Min, LI Wen-Yong, XU Jian-Min. An evaluation method for intelligent transportation systems based on DHGF algorithm[J]. Journal Of GuiLin University Of Electronic Technology, 2007(3):61-64.
- [10] 冯志君.灰色层次分析法在人力资源管理中的应用[J].福建电脑,2009(2):141-142.
FENG Zhi-Jun. The use of gray analytic hierarchy process on human resource management [J]. FuJian computer. 2009(2):141-142.

(上接第62页)

了数据挖掘的原理和方法步骤,并以厦门港主航道的大型船舶AIS信息为例,从空间、时间、概率分布三方面展开对AIS信息中反映船舶实时动态以及周边通航环境状况的船舶航向和航速变化率进行挖掘研究,获得了厦门湾内主航道通航环境空间和时间特性以及航向航速变化率概率分布特点。

利用数据挖掘理论与技术为处理海量AIS信息研究提供了新的思路和方法,为AIS信息的利用开拓了广阔的空间。

参 考 文 献

- [1] 吴兆麟,朱军编著.海上交通工程(第二版)[M].大连:大连海事大学出版社,2004.
- [2] 邵哲平,孙腾达,潘家财,等.基于ECDIS和AIS的船舶综合信息服务系统的开发[J].中国航海,2007(2):30-33.
SHAO Zhe-ping, SUN Teng-da, PAN Jia-cai, et al. Development of the Integrated Vessel Information Service System Based On ECDIS and AIS (In Chinese) [J]. Navigation of China, 2007, 30(2):30-33.
- [3] 国际海事组织.通用船载自动识别系统国际标准汇编[G].袁安存,张淑芳编译.大连:大连海事大学出版社,2005.
- [4] Shao Zheping, Sun Teng-da, Pan Jiakai, Ji Xianbiao. Vessel information service system based on ECDIS and AIS[C]. Proceedings of ICTE 2007, ASCE, 1678-1683.
- [5] 纪贤标,邵哲平,潘家财,等. AIS信息分布式采集系统的开发及关键技术[J].上海海事大学学报.2007,28(3):28-31.
JI Xian-biao, SHAO Zhe-ping, PAN Jia-cai, et al. Development of distributed data collection system of AIS information and key techniques[J]. Journal of Shanghai Maritime University. 2007, 28(3):28-31.
- [6] JI Xian-biao, SHAO Zhe-ping, PAN Jia-cai, et al. A new AIS-based way to conduct OLAP of maritime traffic flow [C]. Proceedings of ICTE 2009, ASCE, 3718-3723.
- [7] Jiawei Han and Micheline Kamber. 数据挖掘概念与技术(第2版)[M].范明,孟小峰译.北京:机械工业出版社,2007.
- [8] D. Hand, H. Mannila and P. Smyth. 数据挖掘原理[M].张银奎,廖丽,宋俊,等译.北京:机械工业出版社,2003.
- [9] 黄燕,吴平. SAS统计分析及应用[M].北京:机械工业出版社,2006.
- [10] 朱世武. SAS编程技术教程[M].北京:清华大学出版社,2007.