

文章编号:1671-1637(2016)05-0142-09

AIS 报文异常动态信息甄别方法

刘兴龙^{1,2}, 初秀民¹, 马 枫¹, 雷进宇^{1,2}

(1. 武汉理工大学 国家水运安全工程技术研究中心, 湖北 武汉 430063;

2. 武汉理工大学 能源与动力工程学院, 湖北 武汉 430063)

摘要:针对船舶自动识别系统(AIS)报文中的异常动态信息,提出一种基于概率推理的包括先验知识提取、证据建模、证据合成与权重系数优化4个步骤的识别方法,运用似然度建模方法将经过人工识别的AIS数据中的速度、航向角和轨迹位置信息转化为0~1之间的证据信度,并用证据推理(ER)规则合成,以验证过的AIS数据作为输入,采用非线性优化方法修正证据权重系数,利用武汉天兴洲大桥水域轮渡与武桥水域货船的AIS数据进行验证试验。试验结果表明:在优化权重系数下武汉天兴洲大桥水域轮渡的正确数据、错误数据、总体数据识别准确率分别为91.67%、97.62%、92.63%;以总体偏差最小为目标时,武桥水域货船的正确数据、错误数据、总体数据识别准确率分别为91.79%、89.87%、91.65%;以正确数据偏差最小为目标时,武桥水域货船的正确数据、错误数据、总体数据识别准确率分别为93.18%、49.95%、90.03%。可见,基于ER规则的AIS动态信息甄别方法能针对不同的优化目标灵活调整证据权重系数,具有接近人工水平的识别准确率。

关键词:交通信息工程;船舶自动识别系统;似然度建模;概率推理;非线性优化;信度分布;数据甄别
中图分类号:U666.1 **文献标志码:**A

Discriminating method of abnormal dynamic information in AIS messages

LIU Xing-long^{1,2}, CHU Xiu-min¹, MA Feng¹, LEI Jin-yu^{1,2}

(1. National Engineering Research Center for Water Transport Safety, Wuhan University of Technology,

Wuhan 430063, Hubei, China; 2. School of Energy and Power Engineering,

Wuhan University of Technology, Wuhan 430063, Hubei, China)

Abstract: Aiming at the abnormal dynamic information in ship automatic identification system (AIS) messages, a recognition approach based on probabilistic inference with four steps including prior knowledge extraction, evidence modeling, evidence combination and weight coefficient optimization was proposed. Likelihood modeling approach was used to transform artificially identified velocity, course angle and track position information included in AIS data to evidence reliability between 0 and 1 that was composed by evidential reasoning (ER) rule. The verified AIS data was regarded as the input, and nonlinear optimization approach was used to modify the weight coefficient of evidence. The AIS data of ferry in Wuhan Tianxingzhou Bridge waters and the cargo ships in Wujiao waters were used to carry out verification test. Test result shows that the recognition accuracies of correct data, incorrect data and total data for ferry in Wuhan

收稿日期:2016-03-01

基金项目:国家自然科学基金项目(61273234, 51479155);湖北省自然科学基金项目(2013CFA007);交通运输部信息化技术研究项目(2013-364-548-200)

作者简介:刘兴龙(1987-),男,湖北松滋人,武汉理工大学工学博士研究生,从事交通信息工程及控制研究。

导师简介:初秀民(1969-),男,吉林通化人,武汉理工大学研究员,工学博士。

Tianxingzhou Bridge waters under optimized weight coefficients are 91.67%, 97.62% and 92.63% respectively. When the minimum total deviation is goal, the recognition accuracies of correct data, incorrect data and total data for cargo ships in Wuqiao waters are 91.79%, 89.87% and 91.65% respectively. When the minimum deviation of correct data is goal, the recognition accuracies of correct data, incorrect data and total data for cargo ships in Wuqiao waters are 93.18%, 49.95% and 90.03% respectively. Obviously, the discriminating method of AIS dynamic information based on ER rule can flexibly adjust the weight coefficient of evidence with different optimized goals, and has the accuracy close to the artificial level. 10 tabs, 6 figs, 25 refs.

Key words: traffic information engineering; ship automatic identification system; likelihood modeling; probabilistic inference; nonlinear optimization; reliability distribution; data discrimination

Author resumes: LIU Xing-long(1987-), male, doctoral student, +86-27-86581899, liuxinglong_its@163.com; CHU Xiu-min(1969-), male, researcher, PhD, +86-27-86581899, chuxm@whut.edu.cn.

0 引 言

水路运输由于运量大、成本低、能耗小、污染少等优势,在中国综合交通运输体系中占有重要地位。航运管理信息化和智能化系统的相关研究一直是国际航运界关注的问题,而获得实时在航船舶交通信息是构建智能航运系统的关键^[1]。船舶自动识别系统(Automatic Identification System, AIS)是目前获取船舶交通信息的重要手段。AIS由岸基(基站)设施和船载设备共同组成,是一种广播式的船-船、船-岸信息自动交互工具,是重要的船舶助航导航设备。船载 AIS 船台在船舶正常航行状态下通常每隔 6~30 s 发送一次 AIS 报文,海事监管部门通过岸基 AIS 基站接收船台发送的 AIS 报文,实现对船舶航行动态的实时跟踪与监控^[2]。AIS 船台发送的报文通常包括:MMSI 编号、船舶类型、船舶在航状态、外形尺寸等静态信息和船舶速度、航向角、经纬度坐标等动态信息。由于 AIS 报文包含的信息容量大,所以 AIS 是分析水上交通行为与风险的重要数据来源^[3-6]。然而, AIS 缺乏完整的信息验证机制,在实际应用中由于定位设备、软件设计等问题,其动态信息包含大量的错误数据^[7-8]。这些错误的信息会干扰海事监管,导致对水面交通态势的误判,同时也会对以 AIS 数据为基础的船舶交通行为分析产生困扰,因此,需要深入研究 AIS 数据正确性识别方法,保障水上交通安全。

目前,在 AIS 报文动态信息甄别方面,马枫等经过对 AIS 信息特征细致的研究,首次提出错误 AIS 数据的甄别问题^[9-10],认为甄别错误 AIS 数据

本质上是建立与人工经验相似的智能算法,并采用 DS 证据理论和改进的 DS_mT 理论建立了剔除错误 AIS 目标的理论框架^[11-12],但该框架只是基本的异常 AIS 数据识别框架,在证据构建方面缺乏严格的推理过程;Zhang 等通过直接删除 AIS 目标轨迹畸变点的方式剔除 AIS 异常数据^[13],这种处理方法显得臆断,且无法处理速度和航向角异常的 AIS 数据;Sang 等采用平滑法删除速度、航向角和轨迹出现异常的 AIS 数据^[14],但对于如何确定合理阈值缺乏严格的探讨;Zhang 等利用 BP 神经网络对远程监控船舶的运动态势进行了估计^[15],利用其自我学习功能对船舶的未来运动态势做了较为精确的预测;Lainiotis 等研究了运动船舶的轨迹实时预测问题^[16],将船舶运动考虑成系统部分未知的预测或估计问题,通过构建动态反复神经网络物理模型,获得比卡尔曼滤波更高的定位精度,而且计算量大大降低。以上模型的成功是以船舶各项参数已知为前提,而 AIS 动态信息中仅仅包含船舶速度、航向角和轨迹位置等少量参数,缺乏足够的参数输入,因此,这些模型不能完全适用于 AIS 动态信息的甄别问题。事实上,船舶在特定水域航行时,其运动特征如速度、航向角和轨迹位置等遵循一定的规律^[17-18],海事监管人员在长期的工作中已经发现并掌握了这些规律,并能够以这些规律为基础准确判断 AIS 动态信息的正确性。然而,面对不断增多的海量 AIS 数据,完全依靠人力无法胜任这项工作,因此,模拟人工辨别 AIS 数据的过程,建立数学模型进而实现自动识别错误 AIS 数据是可行的办法。参考人工识别 AIS 数据的过程可以发现,船舶速度、航向角

和轨迹位置的普遍规律是判断 AIS 数据真实性的重要依据,监管人员通过融合这 3 个证据,得出最终判断结果。这是一个典型的基于先验数据判断后验数据的概率推理模型,因此,可以从概率角度推理 AIS 动态信息的正确性。在概率推理中,常用的模型有贝叶斯推理、贝叶斯网络、DS 证据理论等,其中贝叶斯推理和贝叶斯网络为常用的经典模型^[19-20]。然而,这两者均要求证据相互独立和互斥,且先验概率稳定,而 AIS 船台发出错误数据由硬件故障、软件设计、无线电环境等多方面因素引起,其先验概率并不稳定^[21]。DS 证据理论是对贝叶斯理论的拓展,通过将识别框架扩展到超幂集,能够处理证据间不独立的情况,满足弱概率条件下的证据推理。但是,DS 证据理论自诞生以来就面临大量悖论和质疑,例如 0-1 悖论和证据失效悖论^[22]。对此, Yang 等认为 DS 证据理论中证据间的冲突是由假设证据完全可靠而引起的,为此将证据间的冲突系数分配给全局,并引入证据可靠性系数和权重系数,形成了新的证据推理(Evidential Reasoning, ER)规则^[23]。同时,针对先验概率不稳定情况下的证据信度建模问题, Yang 等对贝叶斯推理进行了拓展,提出了新似然度推理方法^[24],对 DS 证据理论的改进,使得最新的 ER 规则完全适用于 AIS 动态信息甄别问题。

基于以上分析,本文应用 ER 规则提出新的 AIS 动态信息甄别方法,从先验知识获取、证据建模、证据合成和证据权重系数优化 4 个步骤讲述模型的构建方式,然后结合 2 个实例分析不同优化目标下该方法的识别效果。

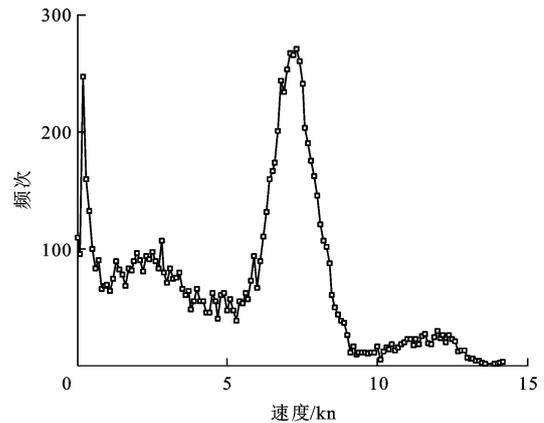
1 模型构建

1.1 先验知识获取

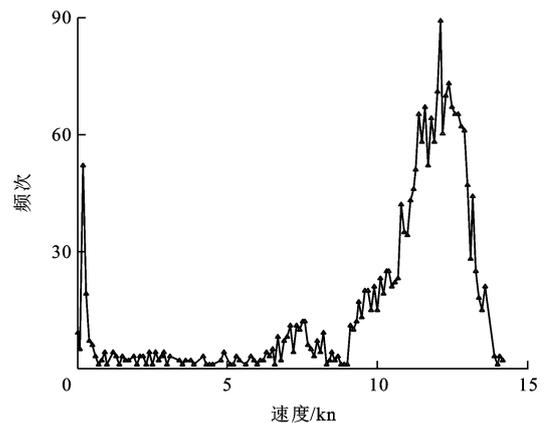
以长江武汉航段天兴洲大桥水域为例,展示该水域轮渡的运动规律。通过海事人员人工识别,分离了该水域 2014 年 5 月 28 日至 2014 年 9 月 23 日期间, MMSI 编号为 413932556 的轮渡在航行途中播发的部分 AIS 数据,有正确数据 10 004 个,错误数据 2 159 个。对这些 AIS 数据进行统计,分别分析正确 AIS 数据和错误 AIS 数据的速度、航向角和轨迹位置分布。

AIS 数据中的速度精度为 0.1 kn, AIS 数据的速度-频次分布见图 1。可以看出,天兴洲大桥水域轮渡的正确 AIS 数据和错误 AIS 数据的速度-频次分布明显不同。

将 AIS 数据中的航向角精度设定为 2° , 统计得



(a) 轮渡正确AIS数据



(b) 轮渡错误AIS数据

图 1 轮渡 AIS 数据的速度-频次分布

Fig. 1 Speed-frequency distributions of ferry AIS data

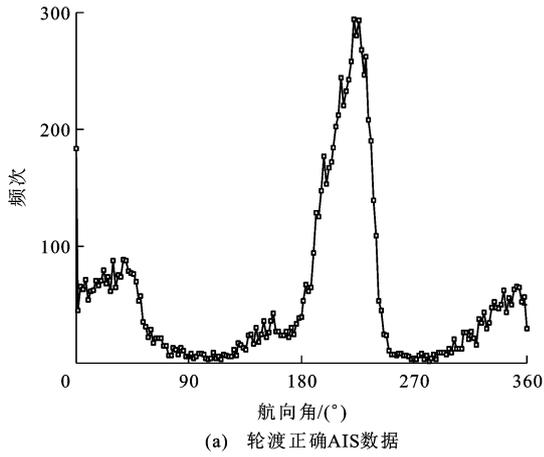
出 AIS 数据的航向角-频次分布见图 2。

不同于速度和航向角, AIS 数据中的轨迹位置以经纬度坐标的形式呈现,是一个二维数组,因此,需要建立二维平面数组,将 AIS 报文中的经纬度坐标映射到二维平面数组中。将研究对象水域以 $0.000 2^\circ$ 的经纬度精度划分为 100×100 的网格,统计 AIS 数据中的轨迹位置数据在网格中的分布。天兴洲大桥水域所处纬度约为北纬 30° , 因此,每个网格对应的尺寸约为 $22 \text{ m} \times 19 \text{ m}$, 相对于船舶外形尺寸该网格划分精度可以满足分析需求。通过对 AIS 数据的经纬度统计,绘制出的轨迹位置-频次分布见图 3。

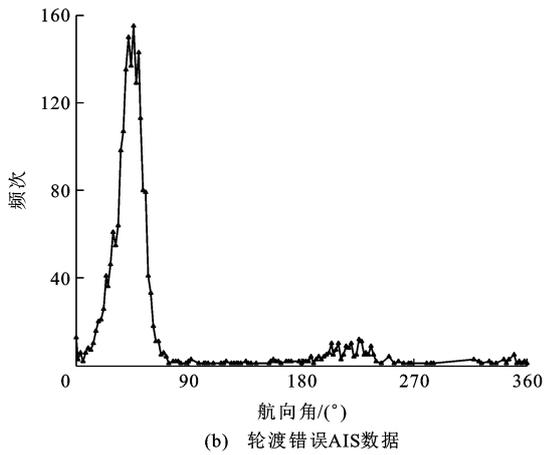
1.2 证据建模

根据速度、航向角和轨迹位置的频次分布,可以发现正确 AIS 数据和错误 AIS 数据的分布存在差异。基于这些差异,任意后验 AIS 数据的速度、航向角和轨迹位置都可以转化为相应的甄别该 AIS 数据正确性的证据。

速度、航向角和轨迹位置证据属于不同维度,难以直接融合,为了将这些证据置于同一个维度下,采



(a) 轮渡正确AIS数据



(b) 轮渡错误AIS数据

图 2 轮渡 AIS 数据的航向角-频次分布

Fig. 2 Course angle-frequency distributions of ferry AIS data

用似然度推理方法将图 1~3 的先验分布转化为 $[0, 1]$ 范围内的证据信度。

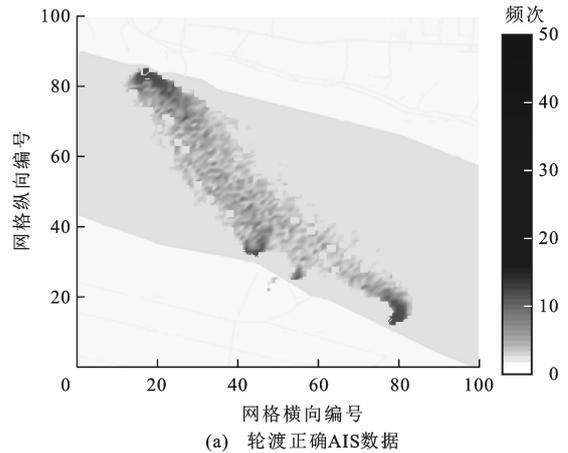
在 AIS 动态信息正确性的判别中存在 2 个互斥的结果:正确数据和错误数据。用 θ_1 代表判定结果为正确数据, θ_2 代表判定结果为错误数据, 则 θ_1 和 θ_2 形成一条 AIS 动态信息的识别框架, 用 $\Theta = \{\theta_1, \theta_2\}$ 表示; Φ 代表空集合, 即判断结果非正确也非错误, 其信度置零; 集合 $\{\theta_1 \cup \theta_2\}$ 代表判断结果为该条 AIS 动态信息可能正确也可能错误。由 Θ 及其子集形成的幂集 $P(\Theta)$ 构成了 AIS 动态信息识别的所有答案, 其形式为

$$P(\Theta) = \{ \Phi, \{ \theta_1 \}, \{ \theta_2 \}, \{ \theta_1 \cup \theta_2 \} \} \quad (1)$$

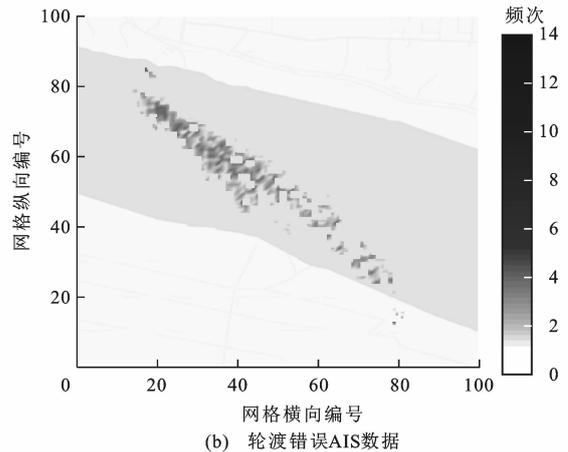
证据建模的核心在于对速度、航向角和轨迹位置 3 个不同的证据分别建立 $P(\Theta) \rightarrow [0, 1]$ 的映射函数, 满足的关系式为

$$\begin{cases} p(\Phi) = 0 \\ \sum_{\theta \in \Theta} p(\theta) = 1 \end{cases} \quad (2)$$

式中: $p(\theta)$ 代表判断结果 θ 发生的概率。



(a) 轮渡正确AIS数据



(b) 轮渡错误AIS数据

图 3 轮渡 AIS 数据的轨迹位置-频次分布

Fig. 3 Track position-frequency distributions of ferry AIS data

速度为 8 kn 的 AIS 数据在图 1(a) 中出现的频次为 107, 出现概率为 $107/10\ 004$; 在图 1(b) 中出现的频次为 7, 出现概率为 $7/2\ 159$ 。从速度来看此 AIS 数据为正确数据的可能性高于为错误数据的可能性, 可知对于任意速度的 AIS 数据为正确还是错误的概率就与图 1 中的速度-频次分布关联起来。

表 1 为样本的频次分布与样本总数。 $f_{i,j}$ 为类别为 j 的第 i 个证据的频次分布, $i=1, 2, 3$ 分别代表速度、航向角、轨迹位置证据, $j=0, 1, 2$ 分别代表数据错误、正确、正确性未知, 本文中无正确性未知的样本, Q_j 为类别为 j 的样本总数。

表 1 样本的频次分布与样本总数

Tab. 1 Frequency distribution of samples and sample total numbers

类别	第 1 个证据	第 2 个证据	第 3 个证据	样本总数
错误	$f_{1,0}$	$f_{2,0}$	$f_{3,0}$	Q_0
正确	$f_{1,1}$	$f_{2,1}$	$f_{3,1}$	Q_1
未知	$f_{1,2}$	$f_{2,2}$	$f_{3,2}$	Q_2

根据表 1, 对于 AIS 数据中速度、航向角、轨迹位置的任意值, 其似然度分布为

$$l_{i,j} = \frac{f_{i,j}}{Q_j} \quad (3)$$

式中： $l_{i,j}$ 为类别为 j 的第 i 个证据的似然度分布。

表 2 为样本的似然度分布。类别为 j 的第 i 个证据的信度分布 $p_{i,j}$ 可以采用归一化的似然度表示为

$$p_{i,j} = \frac{l_{i,j}}{\sum_{j=0}^2 l_{i,j}} \quad (4)$$

表 2 样本的似然度分布

Tab. 2 Likelihood distribution of sample

类别	第 1 个证据	第 2 个证据	第 3 个证据
错误	$l_{1,0}$	$l_{2,0}$	$l_{3,0}$
正确	$l_{1,1}$	$l_{2,1}$	$l_{3,1}$
未知	$l_{1,2}$	$l_{2,2}$	$l_{3,2}$

以 $\{p_{i,0}, p_{i,1}, p_{i,2}\}$ 表示的信度分布代表着 AIS 动态信息中速度、航向角和轨迹位置与判断结果之间的概率关系, 这样将 AIS 动态信息中的速度、航向角和轨迹位置与信度分布建立了映射关系, 并形成判断 AIS 动态信息正确性的证据。

以表 3 的 1 条轮渡 AIS 信息为例, 展示证据建模方法。由表 3 可知, 该轮渡 AIS 数据的速度为 9.9 kn, 该速度在图 1(a)、(b) 中出现的频次分别为 11、21, 因此, 该速度对应于正确 AIS 数据和错误 AIS 数据的似然度分别为

$$l_{1,1} = \frac{f_{1,1}}{Q_1} = \frac{11}{10\ 004} = 0.001\ 10 \quad (5)$$

$$l_{1,0} = \frac{f_{1,0}}{Q_0} = \frac{21}{2\ 159} = 0.009\ 73 \quad (6)$$

表 3 一条轮渡 AIS 信息

Tab. 3 An AIS information of ferry

经度/(°)	纬度/(°)	速度/kn	航向角/(°)
114.418 85	30.662 54	9.9	34

根据式(4), 从速度角度来看该条 AIS 动态信息正确与错误的归一化似然度分别为

$$p_{1,1} = \frac{l_{1,1}}{l_{1,0} + l_{1,1}} = 0.101\ 6 \quad (7)$$

$$p_{1,0} = \frac{l_{1,0}}{l_{1,0} + l_{1,1}} = 0.898\ 4 \quad (8)$$

由于没有不明确的判断, 因此 $p_{1,2} = 0$ 。对于任何 AIS 数据呈现的速度, 均可以根据图 1 求取其为正确或错误 AIS 数据的似然度。这样, AIS 数据的速度转化为判断 AIS 数据正确性的证据。如果某一速度在图 1 中的频次均为零, 即 $p_{1,0}$ 、 $p_{1,1}$ 均等于零, 则此时出现对 AIS 数据正确性的模糊判断, 即 $p_{1,2} = 1$ 。

从图 2 可知, 表 3 航向角出现的频次分别为 75、46, 对应的似然度分别为

$$l_{2,1} = \frac{f_{2,1}}{Q_1} = \frac{75}{10\ 004} = 0.007\ 50 \quad (9)$$

$$l_{2,0} = \frac{f_{2,0}}{Q_0} = \frac{64}{2\ 159} = 0.029\ 64 \quad (10)$$

根据式(4), 从航向角角度来看该条 AIS 动态信息为正确与错误的归一化似然度分别为

$$p_{2,1} = \frac{l_{2,1}}{l_{2,0} + l_{2,1}} = 0.201\ 9 \quad (11)$$

$$p_{2,0} = \frac{l_{2,0}}{l_{2,0} + l_{2,1}} = 0.798\ 1 \quad (12)$$

由于无不明确的判断, 因此, $p_{2,2} = 0$ 。

从图 3 可知, 表 3 经纬度对应出现的频次为 0, 此时出现模糊判断, 即 $p_{3,0} = 0$, $p_{3,1} = 0$, $p_{3,2} = 1$ 。

可见, 通过似然度推理任意 AIS 数据的速度、航向角、轨迹位置都能够转化为 $[0, 1]$ 范围内的信度, 作为判别 AIS 数据正确性的证据。

1.3 证据合成

将 AIS 动态信息中的速度、航向角和轨迹位置转化为可参与证据合成的信度后, 即可对这 3 个证据融合, 获取最终的概率推理结果。在证据处理上 ER 规则相对 DS 证据理论的改进在于, ER 规则引入了证据可靠性系数和权重系数来表征证据的可靠性和重要性, 可靠性是证据能够提供正确判断结果的能力, 权重是证据在合成中出于优化结果的考量决策人主观对证据重要性的评估。用于合成的证据由信度、可靠性和权重三者共同描述, 表达式为

$$\tilde{m}_{\theta,i} = \begin{cases} 0 \\ \frac{\omega_i p_{\theta,i}}{1 + \omega_i - r_i} \\ \frac{1 - r_i}{1 + \omega_i - r_i} \end{cases} \quad (13)$$

式中： $\tilde{m}_{\theta,i}$ 为综合考虑可靠性和权重后由第 i 个证据得出的对判断结果 θ 的信度； ω_i 、 r_i 分别为第 i 个证据的权重系数和可靠性系数； $p_{\theta,i}$ 为第 i 个证据对判断结果 θ 的信度。

用于 ER 合成的证据的最终表现形式 m_i 为

$$m_i = \left\{ (\theta, \tilde{m}_{\theta,i}), \forall \theta \subseteq \Theta; (P(\Theta), \tilde{m}_{P(\Theta),i}) \right\} \quad (14)$$

在 AIS 动态信息识别中, 速度、航向角和轨迹位置 3 个证据均来源于 AIS 船台本身, 可认为这 3 个证据的可靠性系数相同。由于 AIS 船台可靠性影响因素及其作用机理并不明确, 因此, 借鉴已有研究成果将可靠性系数设置为 0.724 4^[25]。

利用式(14),第 1、2 个证据的合成式 $\hat{m}_{\theta,1,2}$ 与对判断结果 θ 的信度分布 $p_{\theta,1,2}$ 分别为

$$\hat{m}_{\theta,1,2} = [(1 - r_2)\tilde{m}_{\theta,1} + (1 - r_1)\tilde{m}_{\theta,2}] + \sum_{B \cap C = \theta} \tilde{m}_{B,1} \tilde{m}_{C,2} \quad \forall \theta \subseteq \Theta \quad (15)$$

$$p_{\theta,1,2} = \begin{cases} 0 & \theta \subseteq \Phi \\ \frac{\hat{m}_{\theta,1,2}}{\sum_{D \subseteq \Theta} \hat{m}_{D,1,2}} & \theta \subseteq \Theta \end{cases} \quad (16)$$

式中: $B、C、D$ 为幂集 $P(\Theta)$ 中除空集以外的任何元素。

以表 3 的 AIS 动态信息为例,当证据可靠性系数和权重系数均为 0.724 4 时,参与合成的证据信度通过式(13)计算,采用 ER 规则融合速度、航向角、轨迹位置 3 个证据则通过式(15)、(16)实现,汇总结果见表 4。表 4 中, $m_{\theta,1,2}$ 为由第 1、2 个证据合成的归一化信度分布, $\hat{m}_{\theta,1,2,3}$ 、 $m_{\theta,1,2,3}$ 、 $p_{\theta,1,2,3}$ 分别为由 3 个证据合成的初始信度分布、归一化信度分布和最终合成结果。

表 4 汇总结果
Tab. 4 Summary result

证据	θ_1	θ_2	$\theta_1 \cup \theta_2$	$P(\Theta)$
$\tilde{m}_{\theta,1}$	0.074	0.651	0.000	0.275
$\tilde{m}_{\theta,2}$	0.146	0.578	0.000	0.276
$\tilde{m}_{\theta,3}$	0.000	0.000	0.724	0.276
$\hat{m}_{\theta,1,2}$	0.071	0.715	0.000	0.076
$m_{\theta,1,2}$	0.083	0.829	0.000	0.088
$\hat{m}_{\theta,1,2,3}$	0.083	0.829	0.064	0.024
$m_{\theta,1,2,3}$	0.083	0.829	0.064	0.024
$p_{\theta,1,2,3}$	0.085	0.850	0.065	

1.4 权重系数优化

可靠性和权重是评估证据质量的参数。证据的权重代表着决策者对证据重要性的评估,不同于可靠性系数,权重系数难以准确客观地评估。证据权重系数的确定取决于设定的优化目标,合适的证据权重系数代表着最终合成结果最贴近实际需求,证据权重系数应当使得总体的判断结果最准确,即对于一定数量的验证样本,判断结果与正确值的偏差之和最小。

对于已经验证的 AIS 数据样本,其正确性已知,可以用确定的数字 0 或 1 来表示其信度。用集合 S_a 、 s_b 分别代表错误 AIS 数据和正确 AIS 数据的验证样本,样本数分别为 m 、 n 。对于正确 AIS 数据样本, A_a 为第 a 个样本判断结果 θ_1 出现的概率;对

于错误 AIS 数据样本, B_b 为第 b 个样本判断结果 θ_2 出现的概率。 A_a 和 B_b 由 ER 规则合成得出,对于每个正确 AIS 数据的判断,偏差可以用 $1 - A_a$ 表示,对每个错误 AIS 数据的判断,偏差可以用 $1 - B_b$ 表示。 A_a 和 B_b 由共同的证据权重系数 $w = \{w_1, w_2, w_3\}$ 确定。对于一定数量的验证样本,总体偏差 $\alpha(w)$ 为

$$\alpha(w) = \sum_{a=1}^n (1 - A_a) + \sum_{b=1}^m (1 - B_b) \quad (17)$$

为最小化总体偏差,证据的权重系数 w 应该使得式(17)取值最小,可得

$$w = \operatorname{argmin}[\alpha(w)] \quad (18)$$

式中: $\operatorname{argmin}(\cdot)$ 为最小化函数; $\alpha(w)$ 为连续可导函数, w 可采用 MATLAB 中自带的非线性优化函数求解。

在确定优化目标之前,证据的初始权重系数和可靠性系数应该相等,即 $w = \{0.724\ 4, 0.724\ 4, 0.724\ 4\}$ 。先探讨总体识别偏差最小目标下的权重系数优化问题,以经过验证的武汉天兴洲轮渡的 2 304 个正确 AIS 数据和 354 个错误 AIS 数据为基础优化证据的权重系数。权重系数优化采用 MATLAB 自带的非线性优化函数工具箱,把目标函数在初始值的邻域泰勒展开,初始值可为人为提供的初始猜测,泰勒展开进行到二阶项为止。在结果判定中以 0.5 作为合成置信度的界限,即在正确 AIS 数据的识别中,当正确 AIS 数据的置信度大于 0.5 时认为识别成功,否则认为识别失败。反之,对于错误 AIS 数据,置信度小于 0.5 时认为识别成功,否则认为识别失败。在初始证据权重系数下对样本的识别结果见表 5。

表 5 初始权重系数下轮渡数据样本的识别结果

Tab. 5 Recognition results of ferry data samples with initial weight coefficient

类别	数据个数	识别成功次数	识别失败次数	准确率/%
正确	2 304	2 099	205	91.10
错误	354	322	32	90.96
总体	2 658	2 421	237	91.08

通过验证数据样本的优化,得到总体偏差最小时优化的权重系数为 $\{1, 1, 0\}$ 。可见优化后轨迹位置证据的权重系数变为零,表示该个证据对降低样本总体偏差没有作用,这是因为相对于识别区域的地理位置网格数量,用于拟合的数据量偏少。采用优化后的权重系数所得验证样本的识别结果见表 6。对比表 5、6,可以发现采用优化后的权重系数

表 6 总体偏差最小时轮渡数据样本的识别结果

Tab. 6 Recognition results of ferry data samples with minimum total deviation

类别	数据个数	识别成功次数	识别失败次数	准确率/%
正确	2 304	2 125	179	92.23
错误	354	323	31	91.24
总体	2 658	2 448	210	92.10

对验证 AIS 数据样本的识别准确率提高了 1.02%。

证据的权重系数可以根据决策人的目标来设定,当海事监管人员认为相对于将错误 AIS 数据误判为正确 AIS 数据,将正确 AIS 数据误判为错误 AIS 数据会导致更加严重的监管风险,此时可将正确 AIS 数据误判为错误 AIS 数据可能性最低作为证据权重系数的优化目标。同样以 0.5 作为判别正确与否的阈值,采用非线性优化后的证据权重系数为{1.000 0,0.780 0,0.000 0},所得验证样本的识别结果见表 7。对比表 6、7,可以发现正确 AIS 数据的识别准确率提高了 3.17%。可见,当分离出有效数据样本时,针对不同的目标采用本文提出的方法可以灵活设置模型参数从而使得模型的结果最优。

表 7 单一偏差最小时轮渡数据样本的识别结果

Tab. 7 Recognition results of ferry data samples with minimum single deviation

类别	数据个数	识别成功次数	识别失败次数	准确率/%
正确	2 304	2 198	106	95.40
错误	354	227	127	64.12
总体	2 658	2 425	233	91.23

2 模型验证与实例分析

为验证算法的有效性,选取 2014 年 9 月 24 日至 2014 年 10 月 22 日期间武汉天兴洲轮渡的 AIS 数据对识别结果进行检验,数据共包含 2 617 个正确 AIS 数据和 504 个错误 AIS 数据。优化权重系数{1,1,0}下的识别结果见表 8。可见,基于 ER 规则的 AIS 动态信息甄别模型对 3 121 个轮渡 AIS 数据的识别准确率为 92.63%,具有很高的准确性。

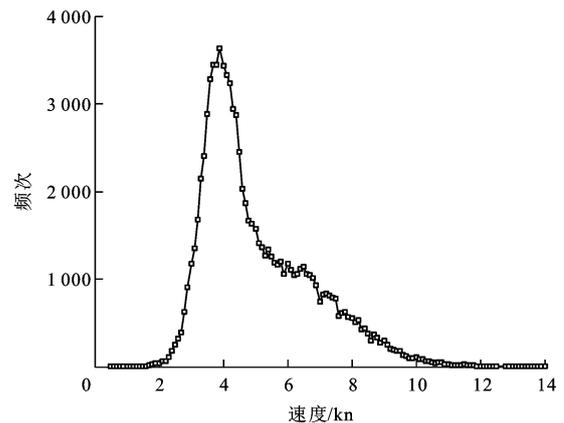
表 8 优化权重系数下轮渡数据样本的识别结果

Tab. 8 Recognition results of ferry data samples with optimized weight coefficient

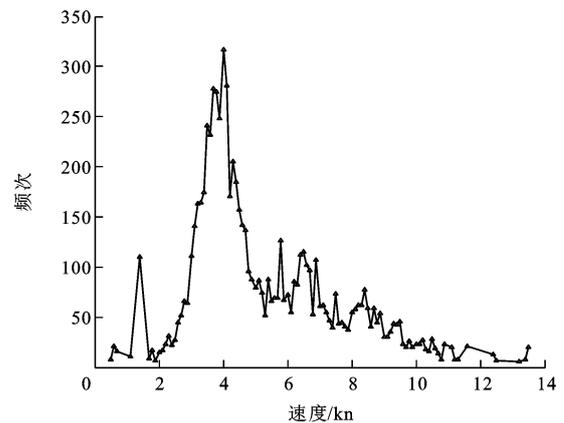
类别	数据个数	识别成功次数	识别失败次数	准确率/%
正确	2 617	2 399	218	91.67
错误	504	492	12	97.62
总体	3 121	2 891	230	92.63

采用同样的方法,通过先验知识获取、证据建模、证据合成和权重系数优化对 2014 年 5 月 28 日至 2014 年 6 月 23 日期间长江中游武桥水域的

2 472 艘货船总计 99 155 个 AIS 数据进行了识别,正确 AIS 数据 91 435 个,错误 AIS 数据 7 720 个。数据的速度、航向角和轨迹位置分布分别见图 4~6。以 2014 年 6 月 24 日至 2014 年 7 月 3 日期间的 25 424 个正确 AIS 数据和 2 089 个错误 AIS 数据作为输入,优化权重系数取值。在初始权重系数{0.724 4,0.724 4,0.724 4}下总体识别准确率为 87.23%。当以总体偏差最小为优化目标时,最优权重系数为{0.923 4,1.000 0,0.024 4},此时总体识别准确率为 91.17%。当以正确 AIS 数据识别偏差最小为优化目标时,最优权重系数为{0.851 4,1.000 0,0.000 0},此时正确 AIS 数据的识别准确率为 93.16%,总体识别准确率为 89.23%。



(a) 货船正确AIS数据

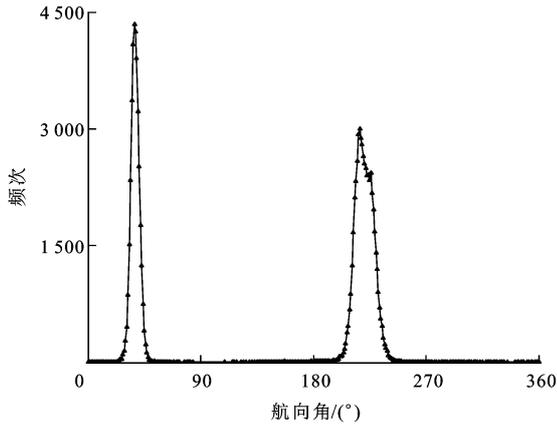


(b) 货船错误AIS数据

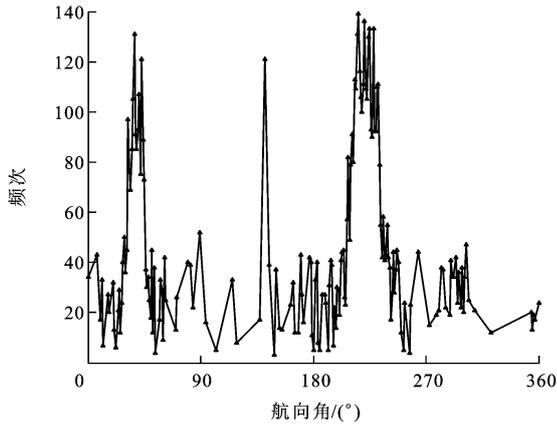
图 4 货船 AIS 数据的速度-频次分布

Fig. 4 Speed-frequency distributions of AIS data for cargo ships

为验证模型的准确性,另取 2014 年 7 月 4 日至 2014 年 7 月 13 日期间的 23 612 个正确 AIS 数据和 1 856 个错误 AIS 数据作为验证样本。当以总体识别误差最小为优化目标时,识别结果见表 9。当以正确 AIS 数据识别偏差最小为优化目标时,识别结果见表 10。由表 9、10 可知,根据不同的优化目标设定证据权重系数可以使得判定结果更贴近实际需求。



(a) 货船正确AIS数据



(b) 货船错误AIS数据

图 5 货船 AIS 数据的航向角-频次分布

Fig. 5 Course angle-frequency distributions of AIS data for cargo ships

表 9 总体偏差最小时货船 AIS 数据的识别结果

Tab. 9 Recognition results of AIS data for cargo ships with minimum total deviation

类别	数据个数	识别成功次数	识别失败次数	准确率/%
正确	23 612	21 673	1 939	91.79
错误	1 856	1 668	188	89.87
总体	25 468	23 341	2 127	91.65

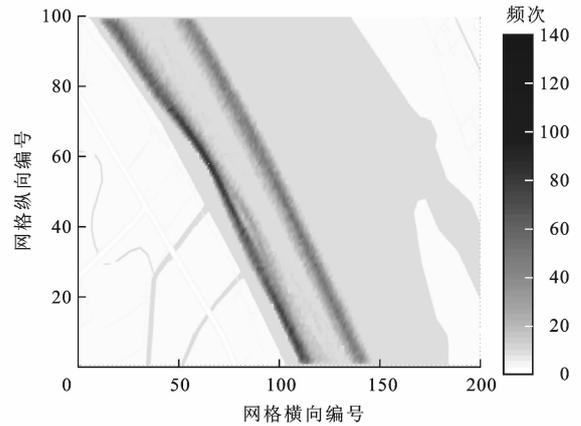
表 10 单一偏差最小时货船 AIS 数据的识别结果

Tab. 10 Recognition results of AIS data for cargo ships with minimum single deviation

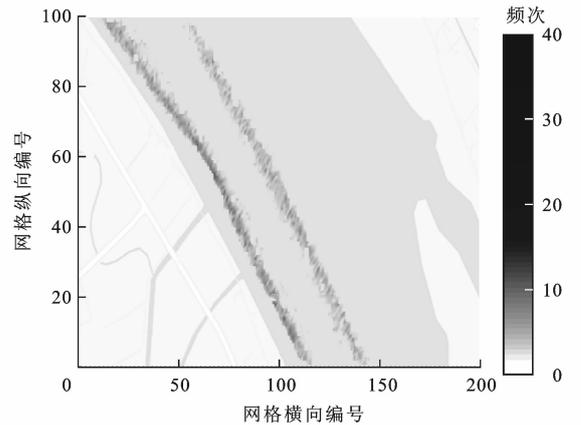
类别	数据个数	识别成功次数	识别失败次数	准确率/%
正确	23 612	22 001	1 611	93.18
错误	1 856	927	929	49.95
总体	25 468	22 928	2 540	90.03

3 结 语

(1)提出了一种基于 ER 规则的 AIS 动态信息甄别方法,具有接近于人工的识别准确率,可代替人工在线识别 AIS 数据,仅以 AIS 数据内含的速度、航向角和轨迹位置作为判断依据即可准确判断数据



(a) 货船正确AIS数据



(b) 货船错误AIS数据

图 6 货船 AIS 数据的轨迹位置-频次分布

Fig. 6 Track position-frequency distributions of AIS data for cargo ships

的正确性。

(2)证据建模采用似然度推理模型,具有不依赖于 AIS 数据全局先验分布的优点。通过非线性优化方法对模型中的证据权重系数进行优化,能够提高数据的识别准确率,且能针对不同的优化目标灵活调整证据权重系数。

(3)由于 AIS 动态信息出现错误的机理尚未弄清,导致证据的可靠性系数取值缺乏严格推理,需要深入探讨证据可靠性系数的赋值机理。本文以经过验证的大量 AIS 数据为基础,数据质量的好坏决定着模型的实际运用效果,因此,在使用模型前需要首先对水域内正确和错误 AIS 动态信息数据进行有效的取样分析。

参考文献:

References :

[1] 孙 星,吴 勇,初秀民. 船-标-岸协同下智能长江航运及其发展展望[J]. 交通信息与安全,2010,28(6):48-52,56.
SUN Xing, WU Yong, CHU Xiu-min. Intelligent Yangtze River shipping and its prospects based on coordination of ship-

- mark-bank[J]. *Journal of Transport Information and Safety*, 2010, 28(6): 48-52, 56. (in Chinese)
- [2] 严新平, 马 枫, 初秀民, 等. 长江船舶交通流实时采集关键技术研究[J]. *中国航海*, 2010, 33(2): 40-45.
YAN Xin-ping, MA Feng, CHU Xiu-min, et al. Key technology of collecting traffic flow on the Yangtze River in real-time[J]. *Navigation of China*, 2010, 33(2): 40-45. (in Chinese)
- [3] PALLOTTA G, VESPE M, BRYAN K. Vessel pattern knowledge discovery from AIS data: a framework for anomaly detection and route prediction[J]. *Entropy*, 2013, 15(6): 2218-2245.
- [4] MASCARO S, KORB B K, NICHOLSON A E. Learning abnormal vessel behaviour from AIS data with Bayesian networks at two time scales[J]. *Tracks A: Journal of Artists Writings*, 2010(8): 1-34.
- [5] TALAVERA A, AGUASCA R, GALVÁN B, et al. Application of Dempster-Shafer theory for the quantification and propagation of the uncertainty caused by the use of AIS data[J]. *Reliability Engineering and System Safety*, 2013, 111: 95-105.
- [6] AARSÆTHER K G, MOAN T. Computer vision and ship traffic analysis: inferring maneuver patterns from the automatic identification system[J]. *TransNav: International Journal on Marine Navigation and Safety of Sea Transportation*, 2010, 4(3): 303-308.
- [7] BALDAUF M, BENEDICT K, MOTZ F. Aspects of technical reliability of navigation systems and human element in case of collision avoidance[C]//RIN. *Proceedings of the Navigation Conference and Exhibition 2008*. London: RIN, 2008: 1-11.
- [8] 朱飞祥, 张英俊, 高宗江. 基于数据挖掘的船舶行为研究[J]. *中国航海*, 2012, 35(2): 50-54.
ZHU Fei-xiang, ZHANG Ying-jun, GAO Zong-jiang. Research on ship behaviors based on data mining[J]. *Navigation of China*, 2012, 35(2): 50-54. (in Chinese)
- [9] 马 枫, 初秀民, 严新平. AIS 基站短消息特性[J]. *交通运输工程学报*, 2012, 12(6): 111-118.
MA Feng, CHU Xiu-min, YAN Xin-ping. Short message characteristics of AIS base stations[J]. *Journal of Traffic and Transportation Engineering*, 2012, 12(6): 111-118. (in Chinese)
- [10] 马 枫, 严新平, 初秀民, 等. 船舶自动识别系统信号失效与场强的相关性[J]. *大连海事大学学报*, 2011, 37(3): 111-114.
MA Feng, YAN Xin-ping, CHU Xiu-min, et al. Correlation between signal failure and field strength in automatic identify system[J]. *Journal of Dalian Maritime University*, 2011, 37(3): 111-114. (in Chinese)
- [11] MA Feng, CHU Xiu-min, YAN Xin-ping, et al. Error distinguish of AIS based on evidence combination[J]. *Journal of Theoretical and Applied Information Technology*, 2012, 45(1): 83-90.
- [12] MA Feng, CHU Xiu-min, LIU Chen-guang. The error distinguishing of automatic identification system based on improved evidence similarity[C]//IEEE. *2nd International Conference on Transportation Information and Safety*. New York: IEEE, 2013: 715-722.
- [13] ZHANG Wen-juan, WU Qing, SANG Ling-zhi, et al. Denoising method of inland AIS information based on vessel track[C]//IEEE. *11th International Symposium on Distributed Computing and Applications to Business, Engineering and Science*. New York: IEEE, 2012: 358-361.
- [14] SANG Ling-zhi, WALL A, MAO Zhe, et al. A novel method for restoring the trajectory of the inland waterway ship by using AIS data[J]. *Ocean Engineering*, 2015, 110: 183-194.
- [15] ZHANG Yue-wen, CUI Wen-bin, WU Gui-tao, et al. A shipboard remote surveillance system improved by BP neural networks[J]. *Navigation of China*, 2009, 32(2): 14-19.
- [16] LAINIOTIS D G, PLATANIOTIS K N, MENON D, et al. Neural network application to ship position estimation[C]//IEEE. *OCEANS Conference 1993*. New York: IEEE, 1993: 384-389.
- [17] XIAO Fang-liang, LIGTERINGEN H, VAN GULJIK C, et al. Comparison study on AIS data of ship traffic behavior[J]. *Ocean Engineering*, 2015, 95: 84-93.
- [18] RISTIC B, LA SCALA B, MORELANDE M, et al. Statistical analysis of motion patterns in AIS data: anomaly detection and motion prediction[C]//IEEE. *2008 11th International Conference on Information Fusion*. New York: IEEE, 2008: 1-7.
- [19] BATTISTELLO G, ULMKE M, PAPI F, et al. Assessment of vessel route information use in Bayesian non-linear filtering[C]//IEEE. *2012 15th International Conference on Information Fusion*. New York: IEEE, 2012: 447-454.
- [20] MASCARO S, NICHOLSO A, KORB K. Anomaly detection in vessel tracks using Bayesian networks[J]. *International Journal of Approximate Reasoning*, 2014, 55(1): 84-98.
- [21] 马 枫. 内河 AIS 通讯链路可靠性分析与技术研究[D]. 武汉: 武汉理工大学, 2012.
MA Feng. Research on the reliability and technology of AIS for inland waterway [D]. Wuhan: Wuhan University of Technology, 2012. (in Chinese)
- [22] YAGER R R. On the aggregation of prioritized belief structures[J]. *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans*, 1996, 26(6): 708-717.
- [23] YANG Jian-bo, XU Dong-ling. Evidential reasoning rule for evidence combination[J]. *Artificial Intelligence*, 2013, 205: 1-29.
- [24] YANG Jian-bo, XU Dong-ling. A study on generalising Bayesian inference to evidential reasoning[C]//CUZZOLIN F. *Proceedings of the Third International Conference on Belief Functions*. New York: Springer International Publishing, 2014: 180-189.
- [25] LIU Xing-long, HE Wei, CHU Xiu-min, et al. A multi-source information fusion method for error AIS targets identification[J]. *The Open Electrical and Electronic Engineering Journal*, 2014, 8: 729-737.