

基于机器学习的复杂船舶信息数据分类方法

赵世鑫 周宁阔 杨志远 任永坤 宗嘉宁

山东交通学院威海校区, 中国·山东 威海 264200

【摘要】传统的复杂船舶信息数据分类方法已经不能满足现有的分类需求,存在着分类性能差的弊端,因此,需要提出一种基于机器学习的复杂船舶信息数据分类方法。首先,基于机器学习剔除复杂船舶信息数据无关特征,得到符合覆盖特征的交叉数据集,其次,利用粗糙集算法,提取交叉覆盖数据集特征,从而实现复杂船舶信息数据的分类。设计对比实验,验证了所提数据分类方法具备更好的分类性能,有一定的应用价值。

【关键词】机器学习; 船舶信息; 复杂数据; 交叉数据集

引言

在研究复杂船舶信息数据分类时,可以利用粗糙集理论。粗糙集理论主要用于解决某些特定的数学问题。在解决某些数学问题时,相关的研究人员发现,其不需要预先设置数据,即可描述属性之间的关系,生成决策集合。在后续的研究中,研究人员将其融入到机器学习中,应用于各种类型的数据分类,但其必须提前界定相关对象的分类能力。应用粗糙集时,需要将需要被处理的数据划分成几个不同的集合,再根据该集合逐步地筛选出需要选择的数据,因此其具有推进能力。

1 基于机器学习的复杂船舶信息数据分类方法设计

1.1 基于机器学习剔除复杂船舶信息数据无关特征

采集复杂船舶的信息数据,在调查中发现其存在着很多数据交叉覆盖区域,这些区域严重影响了数据特征的处理,因此需要剔除复杂船舶信息数据无关特征。此时,设复杂船舶信息的训练集为(1):

$$S = (S_1, S_2, \dots, S_n) \quad (1)$$

公式(1)中, S 代表传输信息交叉数据集, n 代表训练集后续的数据个数,选取该训练集中的部分样本,将其进行部分信息剔除,可得一个数据集,因此,设计此时在训练集样本范围内的投影函数,如(2)所示。

$$\text{dist}(f, si, sj) = \left| \frac{sif - sjf}{\max f - \min f} \right| \quad (2)$$

如公式(2)所示, sif 、 sjf 均代表样品与此时数据特征的关系数值, f 代表投影函数, si 、 sj 分别代表不同的样品, $\max f$ 则代表特征训练集合的最大值, $\min f$ 代表特征训练集合的最小值。

此时采用机器学习算法在训练集中任意选取一个样本 st 。将训练集中与该样本距离最近并且类别相同的样本记作 R_{near} , 将训练集中与该样本距离接近但是类别不同的样本记作 R_{far} , 则每个特征的比重表示如下(3)所示。

$$W(f) = W(f) - \frac{\text{dis}(f, st, R_{near})}{k} - \frac{\text{dis}(f, st, R_{far})}{k} \quad (3)$$

公式(3)中, $W(f)$ 代表此时的特征权重, $W(f)$ 代表标准的权重数值, k 则代表稳定系数, $\text{dis}(f, st, R_{near})$ 代表此时该训练样本在特征函数中的内部最近投影, $\text{dis}(f, st, R_{far})$ 代表

此时该训练样本在特征函数外部的最近投影。

1.2 粗糙集算法提取交叉覆盖数据集特征

本方法采用粗糙集算法,将此时数据集集合中的数据提取出来,为了避免产生数据缺失,首先需要设定数据集领域,列出此时的提取函数如公式(4)所示。

$$f : s \cdot A \rightarrow V \quad (4)$$

公式(4)中, s 代表样本特征, V 代表与特征相对应的集合,即此时列出的粗糙算法的具体属性,此时需要求出该集合具体占据的领域半径,如公式(5)所示。

$$d = \frac{\int V \cdot \beta}{\sigma^2} \quad (5)$$

公式(5)中, d 代表领域半径,代表此时的设定转换函数,为选取公式。

通过此时的覆盖数据特征,以此为依据进一步筛选数据,剔除此时不符合集合特征的非特征数据,此时的集合称为简约交叉覆盖集合,该集合需要设置此时的参数阈值,从而进一步提取覆盖数据的特征,因此其数据进一步提取公式如下公式(6)

$$S' = \frac{\int A_i \cdot S_{red}}{x} \quad (6)$$

根据上述公式,完成了交叉数据的特征提取,为后续进一步提取复杂船舶信息数据特征提供了数据支持。

2 结束语

复杂船舶的信息数据分类一直是比较难的问题,因此本文提出新的信息数据分类方法,实验证明,提出的分类方法有效地提高了复杂船舶信息数据分类的精度与分类的效率,对船舶的安全航行有重要意义。由于实验过程可能受到某些干扰影响,因此需要在应用中进一步改进。

参考文献:

- [1] 刘雨青,冯俊凯,邢博闻,等.基于机器学习的无人船目标识别系统研究[J].测控技术,2019,38(8):31-36.
- [2] 张乔宇,黄国富,金建海.基于机器学习的船舶阻力预报模型研究[J].舰船科学技术,2019,41(12):6-10.