

基于关联规则挖掘的大规模智能电能表在线核查方法

李佳莹, 邢菁, 杨娴, 王丕适, 黄雪玫, 黄开来

(海南电网有限责任公司, 海南 海口 570203)

摘要:为缩短现场核查时间,提升核查效率,提出基于关联规则挖掘的大规模智能电能表在线核查方法。利用采集终端在线采集不同台区的大规模智能电能表运行数据,传送至核查主站。核查主站利用电气特征提取单元,通过DBSCAN聚类算法提取智能电能表的电气特征;利用关联规则挖掘单元,通过Apriori算法挖掘智能电能表电气特征中包含的关联规则;在线核查单元依据关联规则挖掘结果,输出智能电能表在线核查结果。实验结果表明,该方法利用关联规则挖掘结果,可在线核查智能电能表中的计量故障、电源故障、窃电三种异常情况。

关键词:关联规则挖掘;智能电能表;在线核查方法;Apriori算法;电气特征

中图分类号:TP181;TM933 文献标识码:A 文章编号:1003-7241(2025)03-0100-05

Online Verification Method of Large-scale Intelligent Electricity Meter Based on Association Rule Mining

LI Jia-ying, XING Jing, YANG Xian, WANG Pi-shi, HUANG Xue-mei, HUANG Kai-lai

(Hainan Power Grid Co., Ltd., Haikou 570203 China)

Abstract: The online verification method of large-scale intelligent electricity meters based on association rule mining is studied to shorten verification time and improve the verification efficiency. It uses the acquisition terminal to collect the operation data of large-scale intelligent electricity meters in different stations online and send them to the verification master station. The main check station uses the electrical feature extraction unit to extract the electrical features of smart electricity meters by DBSCAN clustering algorithm. The association rule mining unit is used to mine the association rules contained in the electrical features of smart electricity meters by the Apriori algorithm. The online verification unit outputs the online verification results of smart electricity meters according to the mining results of association rules. The experimental results show that the method use the mining results of association rules can check three kinds of abnormal conditions in smart electricity meters online, namely metering fault, power supply fault and electricity theft.

Keywords: association rule mining; intelligent electricity meter; online verification method; Apriori algorithm; electrical characteristics

0 引言

智能电能表是采集电力信息的重要设备,智能电能表的计量质量,决定了电力企业的供电安全性与供电可靠性^[1]。智能电能表核查方式受到众多研究学者重视,高效的智能电能表核查方式,对于提升智能电能表的计量效率具有重要意义。以往智能电能表核查时,需要在智能电能表的不同运行环境下,加载不同大小的负载^[2]。判断不同负载情况下,智能电能表是否能连续准确计量电量,是否存在计量误差。采用该核查方式核查智能电能表,存在浪费大量人力、物力的问题^[3-4]。精准有效的大规模智能电能表核查方式,是电力计量部门以及用电用户极为关注的重要问题。

目前已有大量研究学者针对智能电能表核查问题进行研究。有学者将智能电能表与标准电能表对比^[5],通过对比结果,核查智能电能表是否存在误差。该方法具有核查时间长,浪费大量人力的缺陷,不适用于大规模智能电能表核查;有学者将深度信念网络应用于智能电能表的数据核查中^[6-7],该方法虽然可以核查智能电能表数据中存在的异常,但是存在运算过程过于复杂,电能表核查实时性较差的缺陷。关联规则挖掘方法可以从海量智能电能表运行数据中,挖掘存在关联的规则,为智能电能表数据的实际应用提供基础。

针对以上研究方法存在的缺陷,研究基于关联规则挖掘的大规模智能电能表在线核查方法,将所挖掘的智能电能表运行数据中包含的关联规则,作为大规模智能电能表在线核查的重要依据,为智能电能表的精准核查提供数据基础。

*基金项目:海南省电网有限责任公司科技项目(HNKJXM20200225)

收稿日期:2024-10-28

1 大规模智能电能表在线核查方法

1.1 大规模智能电能表在线核查总体结构

大规模智能电能表在线核查的总体结构图如图1所示。

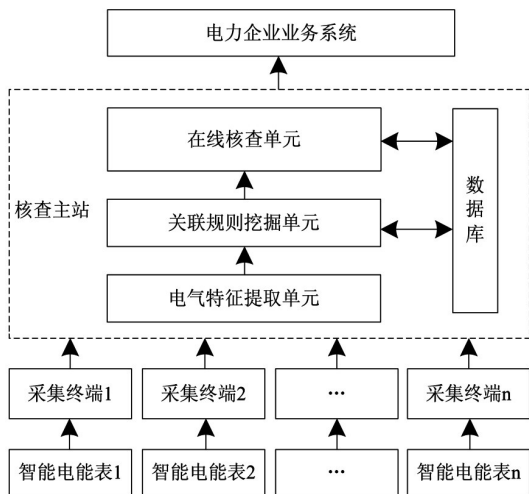


图1 大规模智能电能表在线核查总体结构图

通过图1可以看出,大规模智能电能表在线核查主要包括核查主站、采集终端以及智能电能表三部分。利用采集终端在线采集不同台区的大规模智能电能表运行数据。采集的智能电能表运行数据,传送至核查主站中。核查主站接收大规模智能电能表的运行数据后,利用电气特征提取单元提取智能电能表的电气特征。依据智能电能表的电气特征,利用关联规则挖掘单元,挖掘大规模智能电能表运行数据中包含的关联规则,在线核查单元利用关联规则挖掘结果,核查智能电能表是否存在异常情况。智能电能表和采集终端采用分布式架构部署,核查主站将采集终端所采集的数据合并处理后,为智能电能表在线核查提供决策支持。

采集终端具有数据采集与数据处理的功能,可以视为独立的数据处理单元,具有初步处理以及分类存储智能电能表运行数据的功能,通过分布式处理机制,降低大规模智能电能表在线核查时的运算量。智能电能表通过核查主站,挖掘以及分析大规模智能电能表的运行数据中的关联规则。利用关联规则挖掘结果,核查智能电能表中是否存在异常,精准定位存在异常的智能电能表,将核查结果传递至相应业务系统。核查主站的关联规则挖掘结果以及在线核查结果同步存储至数据库内。后续智能电能表在线核查时,利用历史关联规则挖掘结果,快速挖掘智能电能表的核查结果,提升智能电能表在线核查的实时性。

1.2 提取智能电能表的电气特征

大规模智能电能表在线核查时,提取智能电能表窃电等异常情况时的电气特征信息极为重要。通过采集终端采集智能电能表的负荷数据、用电量等信息后,从所采

集的信息中提取智能电能表的电气特征。利用DBSCAN聚类算法提取大规模智能电能表运行数据中包含的电气特征。DBSCAN聚类算法是一种基于密度的聚类算法,依据样本数量和聚类半径 ϵ ,将待聚类的大规模智能电能表运行数据,划分为核心点、边界点与噪声点三种类型。DBSCAN聚类算法无需确定聚类中心数量,即可实现海量大数据的划分。在噪声干扰情况下,仍然可以精准划分数据,利用数据划分结果提取智能电能表的电气特征。DBSCAN聚类算法利用密度相连和直接密度可达的概念,确定所采集的大规模智能电能表运行数据样本所属的聚类中心。DBSCAN聚类算法运算过程中,智能电能表运行数据样本 p 到达数据样本 q ,组成相同类别,需要满足以下公式:

$$\begin{cases} p \in N(q) \\ |N(q)| \geq MinPts \end{cases} \quad (1)$$

式中, $N(q)$ 与 $MinPts$ 分别表示数据样本 q 的搜寻范围以及半径为 ϵ 的邻域中,数据点的最小数量。满足公式(1)的样本 p 与样本 q 为密度可达的,可以被聚类为相同类别。

利用DBSCAN聚类算法对大规模智能电能表运行数据聚类,设置聚类结果为所提取的智能电能表电气特征。通过聚类提取智能电能表电气特征的流程如下:

- (1) 设定大规模智能电能表运行数据集 S ,聚类半径 ϵ 以及聚类样本数量 $MinPts$;
- (2) 判断输入样本是否满足核心点需求。对于随机样本点 $x_i \in S$,该样本半径为 ϵ 的邻域内,包含对应的样本数量大于 $MinPts$ 时,表示该样本为核心点;
- (3) 输入样本为核心点时,利用公式(1)搜寻该核心点邻域的全部直接密度可达点;
- (4) 重复步骤(2)~(3),直至完成全部样本的判断;
- (5) 将密度可达的样本合并,依据全部核心点邻域内的直接密度可达点^[8-9],搜寻具有最大密度相连点的集合;
- (6) 重复执行步骤(5),直至完成全部核心点邻域的遍历,输出聚类结果。聚类结果即智能电能表运行的电气特征提取结果。

利用DBSCAN聚类算法提取智能电能表的电气特征主要如下:

- (1) 智能电能表倒走。伴随用户用电行为不断累积,电能表用电量不断提升。智能电能表中出现停走或倒走情况时,表示智能电能表存在异常情况,需要进一步判断智能电能表存在窃电行为以及接线错误情况。智能电能表有功总电能低于前一天的智能电能表数据时,表示智能电能表出现倒走情况;智能电能表的随机相电流高于 0.1 A ,同时3日内监测的智能电能表累计电量为0时,表示电能表存在停走故障;
- (2) 智能电能表电流异常。智能电能表电流异常主

要包括电流失流以及潮流反向两种情况。通过智能电能表的三相负载平衡情况^[10],考察智能电能表的电流异常情况。智能电能表电流异常主要包括两种情况:1)电流失流故障。电能表为正常状态时,其电压应高于触发下限。智能电能表中,A相或C相其中一相电流低于0.005倍的B相电流,另一相大于或等于0.05倍的B相电流时,表示存在电流失流情况;2)潮流反向故障。三相电流中随机相有功功率方向存在改变,同时功率高于设定功率时,表示电能表存在潮流反向情况。

(3) 电压异常。智能电能表电压异常主要包括电压越限、电压失压两种异常情况。

除了以上智能电能表故障电气特征外,智能电能表还存在相序异常、开盖记录故障、开关接入状态量变位、恒定磁场干扰等故障。利用DBSCAN聚类算法提取的智能电能表的电气特征,作为大规模智能电能表在线核查的关联规则挖掘的基础,在线核查大规模智能电能表。

1.3 大规模智能电能表在线核查的关联规则挖掘

$I=\{i_1, i_2, \dots, i_m\}$ 表示大规模智能电能表电气特征的数据项集,其中包含数据项 i_k 数量为 n ,子集 $T \subseteq I$ 表示事务,众多事务结合,构成大规模智能电能表运行数据的数据库 D ,其是同智能电能表异常数据挖掘任务具有关联性的事务数据集集合。对数据库 D 中不同数据特征间存在符合电能表异常数据挖掘的关联关系予以呈现的规则即为关联规则。

关联规则挖掘算法即挖掘数据库 D 内包含的关联规则 $A \Rightarrow B, A \subseteq I, B \subseteq I$,且 A 与 B 分别表示前件与后件。 $A \Rightarrow B$ 的关系即为不同电气特征之间的关联规则;用 $\text{Sup}(A \Rightarrow B)$ 与 $\text{Conf}(A \Rightarrow B)$ 分别表示所挖掘关联规则的支持度与置信度。关联规则 $A \Rightarrow B$ 表示某事务中,存在 A 时,具有一定的概率发生 $B, A \subseteq I, B \subseteq I, A \cap B = \emptyset$ 。关联规则挖掘算法中的支持度 $\text{Sup}(A \Rightarrow B)$ 计算公式如下:

$$\text{Sup}(A \Rightarrow B) = \frac{\text{Count}(A \cup B)}{|D|} \times 100\% = P(A \cup B) \quad (2)$$

公式(2)的支持度计算结果,表示事件 A 与事件 B 同时发生的概率。其中, $|D|$ 与 $\text{Count}(A \cup B)$ 分别表示事务组成的数据库大小以及同时包含事件 A 与事件 B 的事务数量。

关联规则挖掘算法中的置信度 $\text{Conf}(A \Rightarrow B)$ 表达式如下:

$$\text{Conf}(A \Rightarrow B) = \frac{\text{Sup}(A \cup B)}{\text{Sup}(A)} \times 100\% = P(B|A) \quad (3)$$

公式(3)的置信度计算结果,表示事件 A 发生时,事件 B 的发生概率。

Sup_{\min} 与 Conf_{\min} 分别表示事务的最小支持度与最小置信度。存在 $\text{Sup}(A \Rightarrow B) \geq \text{Sup}_{\min}$,同时 $\text{Conf}(A \Rightarrow B) \geq \text{Conf}_{\min}$ 时,表示 $A \Rightarrow B$ 为强关联规则;存在 $\text{Sup}(I_k) \geq \text{Sup}_{\min}$ 时,表

示 I_k 为频繁项集 L_k 。该频繁项集则是核查依据。

大规模智能电能表在线核查时,大规模智能电能表电气特征数据项集中频繁项集的挖掘数量与挖掘时间,体现了关联规则挖掘算法的挖掘效率。选取Apriori算法作为大规模智能电能表在线核查的关联规则挖掘算法。Apriori算法挖掘大规模智能电能表电气特征的数据项集关联规则的流程图如图2所示。

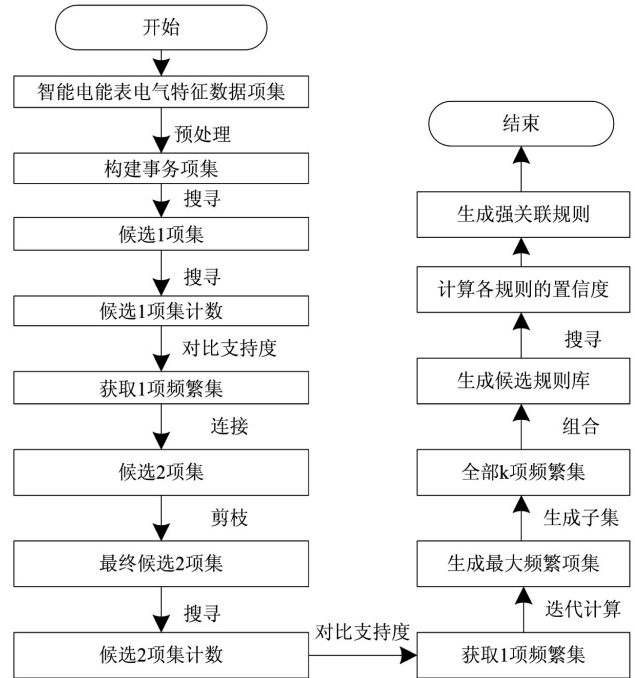


图2 Apriori算法的关联规则挖掘流程图

通过图2可以看出,Apriori算法利用层次迭代思想,以最小支持度作为关联规则挖掘原则,通过连接与剪枝处理,逐层搜索与迭代,直至不再生成满足最小支持度规则的项集。通过以上过程,挖掘频繁项集,生成应用于大规模智能电能表在线核查的强关联规则。

Apriori算法将所获取的频繁项集 L_k ,结合形成候选 k 项集 C_k ,通过剪枝过程,将候选 k 项集 C_k 中的非频繁项集子集删除,获取最终的 k 项频繁集 L'_k ,输出大规模智能电能表在线核查的关联规则与支持度和置信度。Apriori算法适用于大规模智能电能表在线核查的数据挖掘中。利用所挖掘关联规则,输出大规模智能电能表在线核查结果。

2 实例分析

为了验证所研究大规模智能电能表在线核查方法对智能电能表的核查性能,将该方法应用于某电力企业的3个台区共42 528个智能电能表的在线核查中,核查时间段为2020年5月1日-2020年5月31日。将42 528个智能电能表从1-42 528编号,利用采集终端,将智能电能表运行数据实时传送至核查主站。随机选取42 528个智能电能表中的15 484号、25 464号、35 845号智能电能

表,所采集智能电能表运行数据如表1所示。

表1 智能电能表数据采集结果

采集参数	电能表序号		
	15484号	25464号	35845号
编码	15484	25464	35845
电量计量点/kW·h	315 214	285 515	156 482
所属台区	台区1	台区2	台区3
总有功功率	4 854	2 684	4 365
有功功率/W	A相	1 510	945
	B相	2 561	1 025
	C相	952	854
电流有效值/A	A相	215	152
	B相	243	132
	C相	105	95
电压有效值/V	A相	228	224
	B相	229	226
	C相	228	228

给定一个关联规则挖掘的基础事务数据库,其中包含a智能电能表倒走、b智能电能表停走、c电流失流、d潮流反向、e电压失压、f电压越限、g相序异常、h开关接入状态量变位、i开盖记录故障、j恒定磁场故障10项在线核查不合格的因素集合,即项集;包含Z1计量故障、Z2电源故障、Z3窃电三项核查结果组成的事务集。

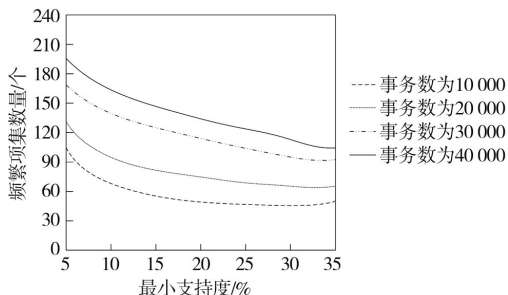


图3 最小支持度对频繁项集数量的影响

统计最小支持度变化时,采用本文方法挖掘关联规则时,生成的频繁项集数量,统计结果如图3所示。通过图3实验结果可以看出,采用本文方法挖掘大规模智能电能表电气特征的数据项集中的关联规则时,伴随所设置最小支持度的不断提升,频繁项集数量逐渐降低。采用本文方法挖掘关联规则时,事务数增加时,频繁项集数量同样有所提升,表明本文方法所采用的关联规则挖掘算法,可以适应大规模智能电能表的海量数据挖掘需求,挖掘性能并未受到数据量提升的影响。进一步分析图3实验结果,伴随最小支持度不断提升,关联规则挖掘的循环迭代次数逐渐减少,获取的频繁项集数量降低。综合考虑关联规则挖掘的运算量以及挖掘结果,设定大规模智能电能表在线核查的最小支持度为10%,最小置信度为90%。

大规模智能电能表在线核查的关联规则挖掘结果如表2所示。通过表2实验结果可以看出,采用本文方法可

以有效挖掘大规模智能电能表在线核查的关联规则。本文方法挖掘的关联规则,为适用于大规模智能电能表在线核查的有效关联规则,可以应用于大规模智能电能表在线核查中。进一步分析表2的关联规则挖掘结果,以a→Z1为例,该关联规则表示大规模智能电能表中,存在某电能表盗走情况时,核查结果为该电能表出现计量故障,与实际智能电能表的运行状态相符。实验样本中,存在该计量故障的实例数共58个,该关联规则的支持度和置信度分别为58.6%以及100%,该关联规则具有较高的支持度和置信度,提升了大规模智能电能表在线核查的性能。

表2 在线核查关联规则挖掘结果

关联规则	实例数量/个	支持度/%	置信度/%
a→Z1	58	58.6	100
{b,c}→Z2	34	49.5	100
{b,d}→Z1	64	38.9	100
{e,g}→Z2	79	33.5	100
{h,i}→Z3	72	29.1	100
{f,g}→Z2	58	28.4	100
{a,c}→Z2	34	24.5	100
{f,j}→Z3	48	23.3	95

依据采用本文方法挖掘的大规模智能电能表在线核查的关联规则,核查实验对象中42 528个智能电能表的运行状态。随机选取其中10个核查判定结果为不合格的智能电能表,核查内容如表3所示。通过表3实验结果可以看出,采用本文方法可以利用所挖掘的关联规则,实现大规模智能电能表的在线核查。本文方法在电力企业中包含大规模智能电能表时,仍然可以实现有效的在线核查,智能电能表在线核查结果,可以作为电力企业调度与管理智能电能表的重要依据。电力企业通过在线核查,及时发现智能电能表存在的故障,快速通知智能电能表管理人员,及时维修以及更换电能表,避免由于电能表故障问题,影响电力企业的正常运营。

表3 智能电能表核查内容

序号	电能表编码	核查时间	核查结果	核查判定结果
1	12 516	2020年5月1日	计量故障	不合格
2	18 564	2020年5月11日	计量故障	不合格
3	21 564	2020年5月8日	电源故障	不合格
4	22 348	2020年5月16日	窃电	不合格
5	25 684	2020年5月7日	电源故障	不合格
6	28 645	2020年5月9日	计量故障	不合格
7	31 526	2020年5月3日	窃电	不合格
8	35 451	2020年5月4日	计量故障	不合格
9	38 645	2020年5月27日	计量故障	不合格
10	40 545	2020年5月19日	电源故障	不合格

3 结束语

采用周期检定等方式核查智能电能表异常时,核查

质量管控水平较低,无法有效监控智能电能表的运行状况。利用关联规则方法挖掘大规模智能电表运行数据中存在的关联规则,实现了大规模智能表的在线核查。关联规则方法可以从海量大规模智能电表运行数据中,挖掘出有价值数据,利用所挖掘数据建立智能电表在线核查的关联规则,为大规模智能电表在线核查提供支撑。智能电表在线核查的实际应用中,伴随大规模智能电表信息数据不断积累,逐步提升核查相关的规则逐渐提升和在线核查的准确率,提升大规模智能电表在线核查的实用性。

参考文献:

- [1] 赵丹枫,黄雁玲,黄冬梅,等.基于AR-TSM的时间序列motif关联规则挖掘方法研究[J].计算机应用研究,2021,38(2):403-408.
- [2] 辛春花,郭艳光,鲁晓波.大型数据库中利用强化学习改进treap的关联规则挖掘算法[J].计算机应用研究,2021,38(1):88-92.
- [3] 杜蜀薇,彭楚宁,徐石明,李天阳.基于贝叶斯层次模型的电能

- 表检定装置在线核查方法[J].电力系统自动化,2018,42(18):177-181.
- [4] 陈昊,乔亚男,刘婧,等.考虑模型病态性的智能电表运行误差分析方法[J].电力建设,2020,41(2):94-100.
- [5] 杨蒂黎,侯兴哲.一种基于安装式标准电能表的智能电表误差在线检测比对方法研究[J].电测与仪表,2022,59(3):181-186.
- [6] 张明泽,栾文鹏,艾欣,等.基于边缘计算的台区短期负荷预测方法[J].电测与仪表,2024,61(4):93-99.
- [7] 肖勇,马喆非,罗鸿轩,等.基于深度信念网络与数据聚合模型的智能电表数据异常检测方法[J].南方电网技术,2021,15(1):99-106.
- [8] 匡伟祥.基于数据挖掘技术的拖拉机发动机故障诊断[J].农机化研究,2025,47(2):244-248.
- [9] 励力帆,虞伟,桑清城.基于数据挖掘的电力系统故障自动诊断方法研究[J].机械制造与自动化,2022,51(2):228-231.
- [10] 曾秀娟,夏群亮,谢慧勤,等.基于TMR的运行监测及故障研判系统研究[J].能源与环保,2022,44(1):222-229.

作者简介:李佳莹(1986-),女,本科,高级工程师,研究方向:电量计量。

(上接第11页)

行驶过程中,队列前方设置人为障碍。图12(a)是机器人的初始位置,三个机器人形成正三角形队形,图12(b)是机器人编队向前直行5 s后的位置,图12(c)-图12(f)分别是机器人编队15 s,30 s,35 s和45 s的位置。可以看到多机器人能够在直行运动和转向运动中避开障碍,维持队形内部的相对位置基本不变。

2.4 实验结果分析

针对6组不同初始位姿的集结编队控制测试,100%能在5 s内完成初始编队的集结,在集结后组成三角形队形。编队协同运动至目标点过程中,人为设置障碍物(障碍物为人),控制算法能在保证基本三角编队队形的情况下避开障碍物,且避开后能继续以稳定编队形式向目标点运动,绕障后回归原规划路线时间在10 s至35 s之间,且100%恢复原有编队队形。考虑到无人车实际车速,该恢复指标较为理想。

通过对实验结果分析可以看出:在多机器人系统执行编队任务形成稳定队形后,当处于实验室环境时,多无人车行驶过程中能够保持编队运行的稳定性和有效性。表明多无人车在实验室环境中编队控制算法具有较高的实时性和准确性,能够实现指定编队运行到目的地的路径规划,在遇到障碍物时能及时绕行障碍物,并在绕行后保持和恢复原有编队运行。

3 结束语

本文以三台无人车作为控制对象,重点讨论了多机器人协同编队控制及其实现问题,针对多机器人队形形

成和队形保持这两个基本问题进行了实验验证,最后搭建多移动机器人软硬件平台,在真实物理环境下实现了多机器人三角形队形的编队运行。本文方法的优越性在于使用的基于ROS2的DDS分布式系统及其通信框架,减少了无人车系统数据交互的开发量,同时提高了多点交互的稳定性,使相应的编队算法有了更高的实时性,加快了遇到障碍物等突发情况的响应速度。此外,提出的基于可变虚拟代理主机的“领导者-追随者”算法,保证时刻有一台“领导者”无人车,从而提高了编队运行的可靠性和灵活性。

参考文献:

- [1] 穆新鹏.多移动机器人协同编队控制策略研究[D].北京:北京邮电大学,2020.
- [2] 王嘉良.基于人工智能技术的机器人目标识别[J].中国战略新兴产业,2019(42):161.
- [3] 范佳.多自主无人地面车辆系统中的路径分配及规划[D].杭州:浙江大学,2011.
- [4] 岑斌斌.多机器人编队的分布式协同控制方法研究[D].南京:南京大学,2015.
- [5] 原魁,李园,房立新.多移动机器人系统研究发展近况[J].自动化学报,2007,33(8):785-794.
- [6] 高志军,颜国正,丁国清,颜德田,陈忠泽.多机器人协调与合作系统的研究现状和发展[J].光学精密工程,2001,9(2):99-103.
- [7] 孙玉娇,杨洪勇,于美妍.基于领航者的多机器人系统编队控制研究[J].鲁东大学学报(自然科学版),2020,36(1):35-39.

作者简介:王荣浩(1985-),男,博士,硕士生导师,研究方向:智能控制理论与应用。