

基于关系代数的高校门户网站数据传输异构数据聚合方法

卜令朵

(中国石油大学(华东)党委宣传部, 山东 青岛 266580)

摘要: 高校门户网站在信息化时代扮演着至关重要的角色,为师生提供了各种服务和信息资源。研究针对高校门户网站的维护问题,提出了一种基于关系代数的异构数据聚合方法。过程中引入非易失性存储器设计了数据传输异构系统控制策略,基于关系代数对数据关联规则计算方法进行了设计,并利用支持向量机完成对网站数据的聚类。实验结果中,研究方法在4处理器主机中数据库大小为64 k时的带宽达到2 034 Mb/s;在数据传输过程中,数据发送最高时延为59 ms。结果说明研究方法能够有效完成高校门户网站数据传输异构数据聚合,能够为网站维护提供参考方案。

关键词: 门户网站; 关系代数; 支持向量机; 数据传输; 聚合

中图分类号: TN919.3; TP311.13 文献标识码: A 文章编号: 1003-7241(2025)03-0079-05

A Heterogeneous Data Aggregation Method for Data Transmission of University Portal Websites Based on Relationship Algebra

BU Ling-duo

(Propaganda Department of the Party Committee of China University of Petroleum (East China), Qingdao 266580 China)

Abstract: University portal websites play a crucial role in the information age, providing various services and information resources for teachers and students. A heterogeneous data aggregation method based on relational algebra is proposed to address the maintenance issues of university portal websites. During the process, non volatile memory is introduced to design a control strategy for heterogeneous data transmission systems. A data association rule calculation method is designed based on relational algebra, and Support Vector Machines are used to cluster website data. In the experimental results, the research method achieves a bandwidth of 2 034 Mb/s on a 4-processor host with a database size of 64 k. During data transmission, the maximum latency for data transmission is 59 ms. The results indicate that the research method can effectively achieve heterogeneous data aggregation for data transmission on university portal websites, and can provide reference solutions for website maintenance.

Keywords: portal website; relational algebra; Support Vector Machine; data transmission; aggregation

0 引言

高校门户网站是大学与多方交流的重要平台。这些网站通常包含了大量异构数据,管理和传输这些异构数据是门户网站维护的重要组成部分^[1-2]。然而,不同数据源之间的数据异构性和数据传输问题使得高校门户网站面临管理和更新上的挑战^[3]。一些平台采用数据库管理系统进行数据管理和传输。尽管这些系统能够提供结构化存储和查询,但在处理异构数据时可能受到数据模式不一致、数据冗余等问题的制约。Aggoune A针对异构数据处理的问题提出了一种智能数据集成手段,利用分层架构和模糊逻辑进行数据集成,结果表明所提方法能提升数据集成的性能^[4]。一些高校门户网站使用人工手动方法来处理数据聚合,进行人工操作容易出现错误,而且不适用于大规模数据传输和异构数据的整合。高光勇

等人针对无线传感器网络的数据聚合问题,提出了一种基于可逆数字水印的方法,使用椭圆曲线进行数据加密,并利用聚合树提升传输效率,结果表明所提方法具有良好的数据处理性能^[5]。进行数据整合时采用关系代数的方法可以更好地处理数据的语义,通过关系运算来对数据进行整合^[5]。关系代数提供了一种抽象的、数学化的数据操作框架,能够更自然地描述数据之间的关系。Vitsaxaki K等人针对大型关系数据集的交互式可视化,提出了一种基于关系代数的方法,使用图表对数据库进行表示,将操作信息进行分组查询,结果表明所提方法能提升交互性能^[6]。在这样的背景下,研究设计基于关系代数的高校门户网站数据传输异构数据聚合方法,以期所高校门户网站编写和维护提供可行的技术参考。

1 高校门户网站数据传输异构数据聚合方法设计

1.1 高校门户网站数据传输异构系统控制策略设计

*基金项目: 山东省自然科学基金项目 (ZR2020MA041)

收稿日期: 2023-11-30

高校门户网站通常包含众多信息源,包括学术、行政、科研、学生管理等各个领域的数据^[7]。网站内的数据不仅包括结构化数据,还包括半结构化和非结构化数据,如文本、图像、视频等,使数据具有更高的复杂性^[8-9]。设计高校门户网站数据传输异构系统控制策略有助于将多源数据整合成一个一致的数据视图,方便网站用户对数据进行更便捷的访问和提取^[10-11]。研究针对高校门户网站数据传输设计异构系统调度策略。调度策略由基本架构模型和管理策略组成,其中基本架构模型包含硬件加速器的工作模型和非易失性存储器的工作模型,管理策略中包含软硬结合分配策略、数据聚合策略和负载均衡分配策略。硬件加速器在运行时,由储存介质获取内存中的数据信息,再将数据传输至加速器的处理单元中进行计算,完成计算后将计算结果发送至主机中。使用硬件加速器进行计算时,任务的完成时间为硬件加速器运行时间和数据传输时间之和。硬件加速器处理数据块的用时如式(1)所示。

$$T_A(s) = \frac{C \times s^\alpha}{A} \quad (1)$$

式(1)中, $T_A(s)$ 代表加速器处理用时; C 是常数系数; s 代表数据块大小; α 代表数据粒度指数系数; A 代表加速器加速比。使用一个数据粒度临界值对加速前后的运行效果进行判断,作为数据分配策略的参考指标。传统的存储器系统使用磁盘作为介质,在读写时不断进行内存同步,磁盘的刷回时间导致了较高的数据传输时延。研究使用非易失性存储介质设计非易失性存储器作为系统的内存介质。使用非易失性存储器作为缓存体系的磁盘写回方式如图1所示。

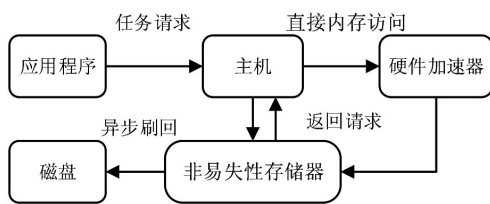


图1 磁盘写回方式

由图1可见,在系统运行时,由网站向主机发起任务请求,主机通过直接内存访问方式将任务发送至硬件加速器中,主机和硬件加速器分别同时将任务信息和处理后数据传输进非易失性存储器中。非易失性存储器向主机返回请求,并在工作空闲时段向磁盘进行异步刷回。整个过程省去了工作时的刷回等待时间。软硬结合分配策略包含软件方式和硬件方式,在不同的应用任务中,软硬件有不同的应用特性,设计软件和硬件的分配方案,如式(2)所示。

$$Se = \begin{cases} H & (process > \gamma \text{ or } blocks > Sa) \\ S & (others) \end{cases} \quad (2)$$

式(2)中, Se 代表所选择的分配方式; H 代表选择硬件调度; γ 和 Sa 代表自适应调度值; S 代表选择主机软件调度; $process$ 代表进程数; $blocks$ 代表操作数据块大小。在使用硬件调度时,若存在其他紧急任务但空闲资源不足时,可以对剩余软件资源进行调用。在使用硬件进行多并发请求任务处理时,若存在较多空闲主机软件资源,则可调用软件资源对部分计算任务进行执行。在运行时,划分一定量的主机资源管理硬件,减少主机的控制资源使用量并提升硬件资源的利用效率。在高校门户网站中存在较多种类数据,不同数据的分布情况具有一定差别,需要对非连续型数据进行聚合以提升通信质量。研究设计数据聚合策略建立在支持向量机和关联规则技术上。负载均衡策略包含一般情况和特殊情况两种不同场景的策略。当所有数据粒度都远小于数据聚合阈值,且数据之间没有相关性时,视为特殊情况场景,其余皆为一般情况。一般情况下使用硬件加速器计算时间和数据传输时间的比值求出数据在硬件加速器中的比值来进行负载分配。特殊情况下使用数据块大小和负载值结合时间和指数系数等参数计算分配比值。

1.2 基于关系代数的数据传输异构数据聚合及路径挖掘方法设计

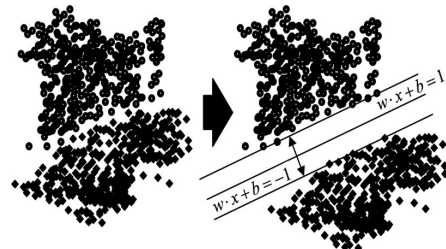


图2 支持向量机运行方式

高校门户网站数据传输异构系统控制策略中的数据聚合策略决定了非连续数据的路径挖掘效果,决定了通信过程中的数据利用率和传输质量^[12]。在进行聚合和路径挖掘时,首先删除数据中的非关键属性数据,获得具有相似样本模式的特征子集,提升数据处理效率。在关系代数理论下的大项集由关系矩阵得出,在基础数据库和项集集合的笛卡尔乘积的子集都属于项集和基础数据库的二元关系。当与基础数据库有关的特征子集所属于的集合为关系数据库子集时,所属的关系数据库子集组成元组对。删除关系数据库子集中的非关联属性,仅对对应属性进行考虑并获得相应的关系矩阵。将第1个属性项集作为1-项集的支持度进行表示。确定1-大项集时使用关系矩阵和最低支持度阈值作为确定元素。在1-大项集的集合基础上获得2-大项集,存在1-1大项集时,将项集元素和阈值进行输入,并规定1不小于2,生成对应的1-项集,完全生成后获得1-大项集。在关系代数技术下获

得非连续型数据的关联规则,提供数据聚合及路径挖掘的依据。支持向量机利用数据集的特点获得不同数据的最佳分类平面,通过最佳分类平面对不同数据进行区分,实现同类别数据的聚类。支持向量机运行方式如图2所示。

由图2可见,在初始的数据分布中,相同类别的数据点分布在一块没有明显边界的区域中。两类不同类别的数据区域在接触边缘中有较为密集的数据点分布情况,且部分不同数据点的位置较为接近。支持向量机插入一个最优分类平面在两类不同数据点群间,并通过平面将两类数据点群向不同方向推离,完成对不同数据的区分。在线性可分情况下的异类中心计算如式(3)所示。

$$o = \frac{1}{n} \sum_{i=1}^n x_i \quad (3)$$

式(3)中, o 代表异类中心; n 代表向量组中向量数; x 代表向量。在非线性可分情况下广义的最优分类超平面计算如式(4)所示。

$$\begin{cases} \varphi(\omega, \xi) = \frac{1}{2} \|\omega\|^2 + \chi \sum_{i=1}^n \xi_i \\ s.t. y_i [\omega \cdot x + b] \geq 1 - \xi_i, 1 = 1, 2, \dots, n \end{cases} \quad (4)$$

式(4)中, $\varphi(\omega, \xi)$ 代表广义最优超平面; ξ_i 代表松弛变量; χ 代表惩罚参数。不同向量间的欧氏距离计算如式(5)所示。

$$d(x, y) = \sqrt{K(x, y) - 2K(x, y) + K(x, y)} \quad (5)$$

式(5)中, d 代表欧式距离; y 代表另一个向量; $K(\cdot)$ 代表核函数。依次计算样本集内样本和异类中心的间距,挑选出待选支持向量。之后遍历样本集并删除重复样本,去除冗余样本。在部分情况下,类中心和支持向量的间距较大,导致支持向量的隶属度较低,所获得的分类超平面不能与最佳分类平面进行匹配。研究对隶属度函数进行修改以调整样本的隶属度值。在包含正样本和负样本时的隶属度函数如式(6)所示。

$$u_i = \begin{cases} \frac{d_i^+ + \phi}{\max d_i^+}, (y = 1, x_i^+ \in X^+) \\ \frac{d_i^- + \phi}{\max d_i^-}, (y = -1, x_i^- \in X^-) \end{cases} \quad (6)$$

式(6)中, u 代表隶属度; ϕ 是一个小数值正数,防止隶属度值为0。为了进一步降低数据聚合时的计算量,研究使用粒子群优化算法对支持向量的数量和体积进行缩减。使用模糊隶属度向量作为粒子群算法的粒子,粒子的移动速度决定了粒子的移动模式。对粒子的位置和速度进行更新,将样本的权重向量进行初始化后根据隶属度调整后的值决定样本取舍。研究所设计的异构数据聚合及路径挖掘流程如图3所示。

由图3可见,在进行异构数据聚合及路径挖掘时,首

先输入训练样本集和关联规则,求出正类中心和负类中心。计算类中心和不同样本的间距,筛选间距小于正类中心和负类中心间距的样本作为待选支持向量。之后计算样本隶属度,生成模糊化待选支持向量集合。再初始化粒子群并将支持向量的权重向量作为粒子,收集粒子位置信息并判别粒子位置。当位置参数大于预设阈值时更改参数为0,获得支持向量集。然后使用粒子适应度调节粒子位置,当更新后适应度值保持不变时结束运行,结合关联规则完成高校门户网站数据传输异构数据聚合及路径挖掘。

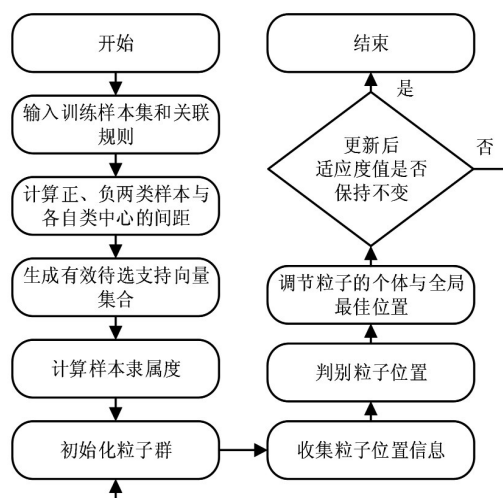
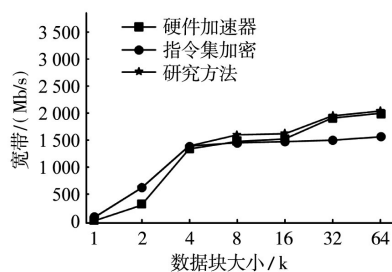
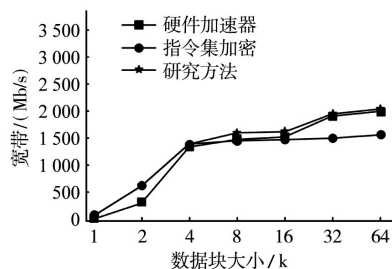


图3 异构数据聚合及路径挖掘流程

2 基于关系代数的高校门户网站数据传输异构数据聚合方法性能测试



(a) 4处理器主机

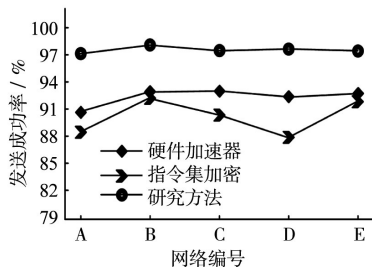


(b) 8处理主机

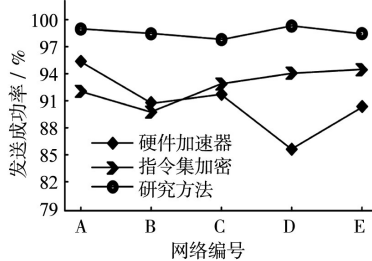
图4 带宽性能测试

为了对研究设计的异构数据聚合方法在高校门户网

站数据传输环境中的有效性进行分析,对研究方法进行性能仿真测试。对方法的带宽性能进行测试,如图4所示。由图4可见,在不同处理器数量的主机中,不同方法的带宽都随着数据块大小的上升而增加。在4处理器主机中,硬件加速器方法在数据块大小为1k时的带宽为43 Mb/s,在数据块大小上升到64 k时的带宽上升到1 977 Mb/s。指令集加密方法在数据块大小为1 k时的带宽为108 Mb/s,在数据块大小上升到64 k时的带宽上升到1 547 Mb/s。研究方法数据块大小为1 k时的带宽为108 Mb/s,在数据块大小上升到64 k时的带宽上升到2 034 Mb/s。在主机处理器数量较少时,研究方法具有更高的带宽值,但相较于其他方法没有较大优势。在8处理器主机中,硬件加速器方法在数据块大小为1 k时的带宽为266 Mb/s,在数据块大小上升到64 k时的带宽上升到1 947 Mb/s。指令集加密方法在数据块大小为1 k时的带宽为513 Mb/s,在数据块大小上升到64 k时的带宽上升到2 772 Mb/s。研究方法数据块大小为1 k时的带宽为513 Mb/s,在数据块大小上升到64 k时的带宽上升到3 476 Mb/s。在主机处理器数量较多时,研究方法具有明显的带宽性能提升,说明研究方法能够有效提升异构系统的系统性能。且研究方法在带宽上升过程中,一直到数据块大小达到64 k时都没有出现性能下降,说明研究方法具有更强的数据块大小适应性能。对网站运行时的数据传输成功率进行测试,如图5所示。



(a) 数据发送

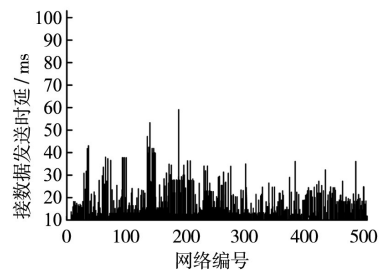


(b) 数据接收

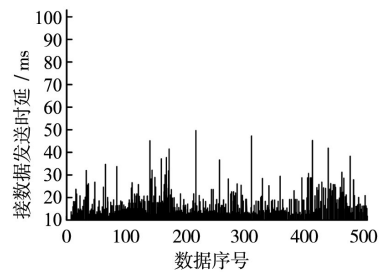
图5 数据传输成功率

由图5可见,不同方法的数据传输成功率在不同网络中具有一定波动。在进行数据发送时,硬件加速器方法在网络编号为C的网络中的发送成功率最高,为92.3%;在网络编号为A的网络中发送成功率最低,为90.6%。

指令集加密方法在网络编号为B的网络中的发送成功率最高,为91.2%;在网络编号为D的网络中发送成功率最低,为87.9%。研究方法在网络编号为B的网络中的发送成功率最高,为98.0%;在网络编号为A的网络中发送成功率最低,为97.1%。在进行数据接收时,硬件加速器方法在网络编号为A的网络中的接收成功率最高,为95.4%;在网络编号为D的网络中接收成功率最低,为85.7%。指令集加密方法在网络编号为E的网络中的接收成功率最高,为94.4%;在网络编号为B的网络中接收成功率最低,为89.8%。研究方法在网络编号为D的网络中的接收成功率最高,为99.3%;在网络编号为C的网络中接收成功率最低,为97.7%。说明研究方法能够有效提升网站的数据传输质量和通信质量。对运行时的数据传输时延进行测试,如图6所示。



(a) 数据发送



(b) 数据接收

图6 数据传输时延测试

由图6可见,在500条数据的传输过程中,数据发送时延和接收时延都在一定范围内波动。在进行数据发送时,约有64%数据的发送时延在20 ms以下;约有2%的数据发送时延在40 ms以上。500条数据中的最高发送时延为59 ms。在进行数据接收时,约有73%数据的接收时延在20 ms以下;约有1%的数据接收时延在40 ms以上。500条数据中的最高接收时延为50 ms。研究方法的最高数据发送接收时延处于可接受范围,且在多数情况下的发送接收时延保持较低值。说明研究方法能够有效在高校门户网站数据传输中完成数据聚合并保障数据的传输速度。

3 结束语

高校门户网站数据的有效聚合和传输是门户网站管

理的一个关键挑战。研究提出了一种基于关系代数的方法以解决高校门户网站数据传输异构数据聚合的问题。过程中使用非易失性存储器作为缓存体系设计了磁盘写回方式,并规划了系统控制策略,之后将支持向量机和关系代数技术进行结合实现对数据的聚类 and 路径挖掘,最后对提出方法的有效性进行了测试。实验结果表明研究方法具有良好的数据传输和聚合性能。但研究仅在有线网络环境中进行测试,后续将在无线网络环境中进行测试并对方法进行优化。

参考文献:

- [1] 薛欢庆,张玲,范广玲.基于关系代数的非连续数据路径挖掘算法[J].计算机仿真,2022,39(11):392-397.
- [2] 夏伟,蔡文婷,刘阳,等.基于联合卡尔曼滤波的配电网多源异构数据融合[J].电力系统保护与控制,2022,50(10):180-187.
- [3] 莫慧凌,郑海峰,高敏,等.基于联邦学习的多源异构数据融合算法[J].计算机研究与发展,2022,59(2):478-487.
- [4] Aggoune A.Intelligent data integration from heterogeneous relational databases containing incomplete and uncertain information[J].Intelligent Data Analysis,2022,26(1):75-99.
- [5] 高光勇,韩婷婷,夏志华.基于可逆数字水印的无线传感器

网络可恢复数据聚合协议[J].计算机科学,2023,50(8):333-341.

- [6] Vitsaxaki K,Ntoa S,Margetis G,Spyratos N.Interactive visual exploration of big relational datasets[J].International Journal of Human-Computer Interaction,2023,39(10):2033-2047.
- [7] 马礼智,唐睿,张睿智,等.基于无线能量传输的物联网数据采集系统中资源分配机制的设计[J].信息与控制,2023,52(2):220-234.
- [8] Wang G,Zhao B,Wu B,et al.Intelligent prediction of slope stability based on visual exploratory data analysis of 77 in situ cases[J].International Journal of Mining Science and Technology,2023,33(1):47-59.
- [9] 陈广智,曾霖,刘伴晨,等.基于Python的电商网站服装数据的爬取与分析[J].计算机技术与发展,2022,32(7):46-51.
- [10] Burian J,Zapletal J,P ászto V.Disaggregator—a tool for the aggregation and disaggregation of spatial data[J].Earth Science Informatics,2022,15(2):1323-1339.
- [11] 余晓帆,朱丽青.基于Flask框架的社交网站数据爬取及分析[J].微型电脑应用,2022,38(3):9-12.
- [12] 方国强,吴雪霁,包森成.基于机器学习的多元异构网络数据安全传输技术[J].自动化技术与应用,2022,41(5):107-149.

作者简介:卜令朵(1986-),女,硕士,编辑,研究方向:网络信息化,多媒体技术。

(上接第51页)

个障碍物,本文方法规划变电站移动机器人巡检路径,路径规划结果如图10所示。图中红色区域为障碍物。由图10可知,机器人在应用本文方法进行路径规划时,不会出现局部极小位置,可有效避开在规划路径的8个障碍物,实验结果表明,本文方法可以规划最短到目的地的距离,并在行进时避开障碍物。

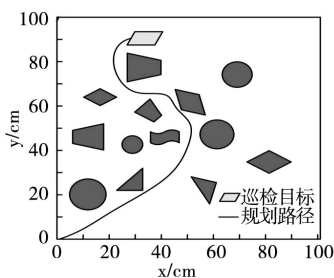


图10 移动机器人巡检路径规划图

3 结束语

本文研究变电站机器人巡检路径规划,通过传感器采集移动机器人的位置信息,利用卡尔曼滤波器融合位置信息,获取机器人当前位置,用人工势场法规划机器人巡检路径。实验证明本文方法可以让机器人避开障碍物,且到达巡检目的地的位置,规划机器人巡检路径最短,让机器人快速地完成巡检任务,在变电站区域巡检规划时,根据巡检任务,安排合适的移动机器人数量。本文

主要研究在机器人巡检时,避开固定的障碍物,未来还需对动态障碍物展开更深入探索。

参考文献:

- [1] 赵建平,于文海,许贤厚,等.变电站智能巡检机器人的拓展应用研究[J].能源与环境,2022,44(1):248-255.
- [2] 吴翻卉,卢杜晓,罗忠辉,等.智能机器人的超声波避障技术研发及应用[J].机床与液压,2019,47(9):46-50,57.
- [3] 徐金雄,王涛,刘军,等.基于双蚁群算法的双机器人路径规划方法[J].机床与液压,2020,48(23):45-48.
- [4] 王军,杨云霄,李莉.基于改进深度强化学习的移动机器人路径规划[J].电子测量技术,2021,44(22):19-24.
- [5] 侯晓晓,涂亚楠,陈晓静,等.基于鲁棒模型预测的无人拖拉机跟踪控制方法分析[J].农机化研究,2025,47(4):242-246.
- [6] 李晨阳,彭程,张振乾,等.融合里程计信息的农业机器人定位与地图构建方法[J].农业工程学报,2021,37(21):16-23.
- [7] 杨彦海,岳靓,杨野.基于超声波法的冻融作用下冷再生混合料路用性能预测研究[J].森林工程,2023,39(1):140-148.
- [8] 王翥,刘春龙,罗清华.超声波传感器特性分析与测试方法的研究[J].郑州大学学报(工学版),2020,41(2):13-18.
- [9] 孙文成,李健,彭宇辉,等.基于样本不均衡和特征优选多源融合的输电线路故障类型辨识[J].电测与仪表,2024,61(12):79-89.
- [10] 彭文正,敖银辉,黄晓涛,等.多传感器信息融合的自动驾驶车辆定位与速度估计[J].传感技术学报,2020,33(8):1140-1148.

作者简介:王高(1987-),男,本科,工程师,研究方向:机电。