

物联网采集缺失信息下的高精度虚拟重构算法

朱婉晋¹
ZHU Wanjin

摘要

传统方法在处理物联网设备采集的数据时,往往难以应对数据缺失的问题。高精度虚拟重构算法能够通过分析现有数据的模式和关联性,填补缺失的信息,从而提高数据集的完整性。为此,文章提出了一种针对物联网采集缺失信息的高精度虚拟重构算法。利用 MapReduce 框架进行数据的并行化预处理,包括数据清洗和异常值剔除,以提升数据采集的整体质量,并有效提取物联网中的缺失信息特征。设计了缺失信息特征识别器,此识别器利用模板匹配和相似度,度量精准识别数据集中的缺失部分。在构建虚拟重构模型时,考虑节点互联能力阈值、链路传输功率总和限制等多重约束条件,通过历史数据训练模型以学习适应不同数据缺失情况。提出了重构排序函数,针对不同物理链路条件采取灵活恢复策略,动态调整重构路径,以确保系统的稳定运行和高效数据传输。依据资源条件灵活选择迁移或分割策略,确保在带宽资源充足时优先完整迁移,资源紧张时则采用动态路径分割,以最优方式恢复物联网数据的完整性与连续性。实验结果表明,设计算法在五个数据集上的请求接受率均高于文献[1]和文献[2]中的算法,特别是在 DS3 数据集上,其请求接受率高达 93.2%,显著优于其他两种算法,这充分证明了文章的设计算法在复杂物联网环境下具有的高精度重构能力。

关键词

物联网;采集缺失信息;虚拟网络链路;高精度;虚拟重构算法

doi: 10.3969/j.issn.1672-9528.2024.11.010

0 引言

物联网在实际应用中常常面临数据采集不完整的问题,尤其是在复杂多变的环境条件下,传感器故障、通信中断或数据传输延迟等因素都可能导致关键信息的缺失。这些缺失信息不仅影响了物联网系统的整体性能,还可能对基于物联网的决策分析产生误导。因此,如何有效地对物联网采集缺失信息进行高精度虚拟重构,成为当前物联网技术研究的热点之一。

近年来,诸多学者对高精度虚拟重构算法进行了研究。在现有研究中,例如,基于深度学习与压缩感知理论的通用图像重构算法^[1],利用深度学习强大的特征提取和表示能力,结合压缩感知理论在信号处理中的优势,实现了对缺失图像数据的有效恢复。该算法通过训练深度学习模型,学习图像数据的内在结构和特征,从而在缺失部分数据的情况下,能够依据已知信息推断出缺失部分的内容。另一方面,偏振光谱多维信息的高精度虚拟重构融合算法^[2],充分利用了偏振光谱数据在不同维度上的信息互补性,通过特征提取、融合和优化等步

骤,实现了对缺失偏振光谱数据的高精度重构。这种算法在光谱分析、遥感探测等领域具有广泛的应用前景,也为物联网数据重构提供了新的视角。多角度 SAR 图像非目标遮挡缺失信息重构^[3],该算法通过利用多角度 SAR 图像的信息互补性,结合先进的图像处理技术,如深度学习和图像融合算法,实现了对缺失信息高度重构,但是仍存在对计算资源有较高要求、模型训练高成本以及对遮挡区域识别有难度等问题。城市轨道交通车辆管控虚拟重构方法设计^[4],该方法设计时兼顾了计算效率,确保了实时处理和重构图像的能力,以满足城市轨道交通对快速响应的需求,但该方法存在对计算资源的高需求、模型训练的高成本等问题。尽管上述四种算法在各自领域取得了显著成果,但在物联网采集缺失信息的高精度虚拟重构方面仍存在诸多挑战,因物联网数据具有多样性、实时性和海量性等特点。

为了解决上述方法存在的问题,需要针对物联网数据的特性,设计一种全新的高精度虚拟重构算法。本文旨在综合现有研究成果,结合物联网数据的特性,提出一种创新的物联网采集缺失信息下的高精度虚拟重构算法,有助于解决物联网数据采集不完整的问题,还将为物联网技术的进一步发展和应用提供有力支持。

1. 山西金融职业学院 山西太原 030008

1 提取物联网缺失信息特征

物联网系统因其分布式传感器和设备的特性,容易受到设备故障、通信中断和环境干扰等因素影响,导致数据流中出现缺失值。这些缺失不仅损害数据的完整性,还可能误导基于这些数据做出的决策。为了多方面提升系统的整体性能和数据质量,需要对物联网缺失信息特征进行提取。

在深入探索物联网缺失信息特征提取的过程中,本研究采用了一种结合多维空间重构与动态模式分析的创新方法。首先,为了高效处理海量的物联网数据,利用 MapReduce 框架对海量的物联网数据进行高效切分与初步处理,确保数据在 Map 阶段得以并行化分析,显著加速了数据处理流程。在 Map 函数的执行过程中,针对每个数据片段执行特征提取的预处理步骤,如数据清洗、异常值剔除等,以确保数据质量,为后续准备数据进入更复杂的分析阶段奠定基础。

随后,在 Reduce 阶段^[5],引入一种基于张量分解的算法。该算法能够深入挖掘物联网数据中的隐藏结构。假设物联网资源信息可以表示为一个多维张量 $\mathbf{X} \in \mathbf{R}^{I \times J \times K}$,其中 I 、 J 、 K 分别表示时间、空间或不同类型的资源维度。

为了重构包含缺失信息的相空间,定义一个重构张量 $\bar{\mathbf{X}}$,其元素的优化求解公式为:

$$\bar{\mathbf{X}} = \text{Reg}(\mathbf{Z})P\|\mathbf{X}\|_F \lambda \quad (1)$$

式中: $\text{Reg}(\mathbf{Z})$ 表示正则化项,用于防止过拟合并促进解的稀疏性或平滑性, P 表示一个二进制矩阵,用于标记缺失位置, F 表示 Frobenius 范数,用于衡量数据拟合的误差, λ 表示正则化系数。

假设在物联网资源信息重构相空间中,任意两点之间的距离若小于一个预设的阈值 θ ,则称这两点为关联点。给定相空间中存在 k 个物联网资源信息点,旨在提取这些点之间的关联特征,通过计算所有满足距离条件(即小于 θ)的点数量,作为关联强度的量化指标,即关联积分。这一过程公式可表示为:

$$C_\theta = \sum_{i=1}^k \sum_{j=i+1}^k I[d(\bar{X}_i, \bar{X}_j)] \quad (2)$$

式中: C_θ 表示在阈值 θ 下的关联积分, X_i 和 X_j 分别表示相空间中的第 i 和第 j 个物联网资源信息点, d 表示两点之间的距离度量函数。

通过利用关联积分来构建物联网资源信息的关联运算,针对缺失信息的特征,特征提取公式为:

$$D = \alpha \frac{\sigma}{C_\theta} \quad (3)$$

式中: α 表示特征系数, σ 表示物联网资源信息维数。

2 识别物联网缺失信息

物联网系统由于涉及大量设备和复杂环境,数据采集过

程中难免会出现缺失或异常的情况。缺失信息如果不被及时识别和处理,可能会在系统分析、决策或预测时引入偏差,影响系统的准确性和可靠性。通过设计缺失信息特征识别器,可以主动识别出这些包含缺失信息的数据点,分析出哪些设备或环节的数据采集存在问题,从而有机会进行后续的数据处理,如数据补全或异常处理,进而提高整个数据集的质量。

将提取到的物联网资源信息特征进行向量化处理,形成特征向量集 $\mathbf{F} = (D_1, D_2, \dots, D_m)$,其中 D_m 代表第 m 个资源信息的特征向量。考虑到缺失信息的特殊性,采用梯度提升树,有效捕捉特征间的非线性关系^[6]。

特别设计一个缺失信息特征识别器,用于识别那些包含缺失信息特征的数据点。该识别器将数据点的特征向量与预设的缺失信息特征模板 T 进行匹配,利用相似度量函数 $S(D_m, T)$ 计算匹配度。这里,引入一个阈值 γ ,当 $S(D_m, T) > \gamma$ 时,认为该数据点包含缺失信息。

为了更具体地描述这一过程,定义公式:

$$S(D_m, T) = \sum_{m=1}^{\sigma} T_m \otimes |w_m| \quad (4)$$

式中: T_m 表示缺失信息特征模板对应的维度值, w_m 表示第 m 个特征的权重,反映了该特征在识别缺失信息时的重要性。

通过遍历整个物联网资源信息集,应用上述缺失信息特征识别器,能够快速且准确地识别出包含缺失信息的数据点。对于识别出的缺失信息,进一步触发相应的处理机制,如数据补全、异常告警或调整数据采集策略,以确保物联网系统的稳定性和可靠性。

3 建立缺失信息重构模型

在数据密集型和依赖数据驱动领域,如物联网、大数据分析、医疗健康监测和金融风险评估等领域,缺失信息可能导致分析准确性下降、决策支持不足、资源分配效率低下以及系统性能受损。为了应对这一挑战,可以采用重构模型。这种模型能够基于现有数据和其他相关信息,对缺失值进行合理估计和填补,从而减少因数据不完整带来的误差,提高分析结果的准确性和可靠性。因此,一旦识别出包含缺失信息的数据点,便可以运用所构建的重构模型来填补数据集中的空白,恢复数据的完整性。

在物联网的庞大架构中,针对复杂多变的采集节点网络 $G=(V, E)$,其中 V 代表散布分布广泛的传感器节点集合,而 E 则映射了这些节点之间错综复杂的通信链路,特别是因故障或干扰导致信息缺失的关键链路^[7]。为了应对这一挑战,构建一种创新的缺失信息重构模型。该模型基于物联网缺失信息状况监测,智能调度虚拟资源与链路,在物理网络架构上实施动态映射策略。通过精准评估物理网络的负载分布,模型能够灵活调整虚拟元素的位置,旨在优化资源利用,平

衡网络负载，显著提升在面临信息缺失时对虚拟网络请求的接纳能力，从而保障物联网系统的稳定运行与高效数据传输。

鉴于物联网中采集节点信息缺失的局限性，节点 G 的互联能力设有阈值，即 g 代表当前连接数，而 $g_{\max}$ 界定其潜力上限。此外，整个物联网体系内，所有链路传输功率总和受限于 $B_{\max}$ ，这一设定确保了网络运行的能效与安全。在构建缺失信息重构模型时，需精细考量这些约束条件：

$$g \leq g_{\max}, \sum_{g=1}^G P \leq B_{\max} \quad (5)$$

式中： P 表示链路实际传输功率。

此时，针对物联网采集信息缺失环境下，将高精度重构模型构建问题转化为多约束条件下的最优求解问题。这要求在保持物联网稳定性的前提下，最大化信息恢复精度^[8]，通过数学建模方法，将实际问题抽象为可解的约束优化任务，公式为：

$$\max \eta = \begin{cases} P_{\min} \leq P \leq P_{\max} \\ g \leq g_{\max} \\ \sum_{g=1}^G P \leq B_{\max} \end{cases} \quad (6)$$

式中： $P_{\min}$ 表示链路传输最小功率， $P_{\max}$ 表示链路传输最大功率。

使用历史数据对模型进行训练，准备一套全面且具有代表性的历史数据集，这些数据应涵盖物联网系统中各种可能的数据缺失情况和正常运行状态，以确保模型能够学习到足够的特征和信息。训练过程通常涉及多次迭代，每次迭代中，模型都会根据输入的历史数据调整其内部参数，以最小化预定义的损失函数，这个过程也被称为模型的优化。

4 资源条件驱动的高精度虚拟链路重构策略

在 t 时刻缺失信息重构模型检测到故障后，首要任务是识别并统计受影响的虚拟物联网群体，随后运用一种排序机制来确立它们的恢复优先级。针对每个受影响的虚拟物联网，恢复策略被设定为优先处理故障情形 2 与 3，紧随其后是故障情形 1 的修复。若某一虚拟物联网的恢复尝试未能成功，则会自动释放其原先占用的物理网络资源，并触发相应的赔偿金流程^[9]。

为科学合理地确定虚拟网络的恢复顺序，本文综合考量了以下三个维度：其一，聚焦于虚拟物理网因故障可能引发的赔偿金额度，作为评估其紧急性的重要指标；其二，考虑赔偿金累积的时间跨度，即赔偿金从产生到解决所需的时间长短，这也直接影响到资源的高效利用与成本控制；其三，则是虚拟物理网自身的恢复复杂度，具体以虚拟物理网中涉

及的故障节点数与链路数之和 n 来量化，直观反映了恢复的难易程度。基于上述三方面考量，本文定义了一个基于缺失信息重构模型的重构排序函数为：

$$H = \frac{\eta \cos G}{n} \quad (7)$$

当 H 值呈现上升趋势时，它间接揭示了 $\eta \cos G$ 或 t 中至少有一方数值显著。这具有双重含义：一方面， $\eta \cos G$ 的值较高意味着物联网 G 涉及的赔偿金规模庞大，反映了较高的经济损失潜力；另一方面， t 值的增大则指向赔偿金累积时间的延长，这不仅增加了财务负担，还提升了同一时间点内多个物联网 G 同时出现问题的可能性，进而加剧了总体赔偿金的数额。这两种情况均不利于虚拟物联网的效益与成本比率。

针对因各种原因（如设备故障、网络拥塞等）导致的虚拟链路数据缺失或故障，通过高精度虚拟重构算法来实现数据的有效恢复和传输链路的优化。尽管虚拟重构模型已经完成了对缺失信息的填补和重构，但这一过程的成功并不意味着数据传输链路能够立即恢复到最优状态或长期保持高效稳定。为此，本研究提出针对不同物理链路条件采取灵活恢复策略，动态调整重构路径，以确保系统的稳定运行和高效数据传输。在缺失信息的高精度虚拟重构算法上，依据资源条件采取不同方案：若物理链路带宽资源充裕，则优先考虑将故障虚拟链路完整迁移至一条独立的物理通道上，确保传输连续性；若面临带宽资源紧张的情况，则启动动态路径分割策略，将故障虚拟链路拆分为多个部分，分别安置于不同的独立物理路径上，以此提升恢复成功率并优化资源利用。具体执行流程如下。

（1）排序与选择：首先，根据故障虚拟链路的带宽需求进行降序排列，逐一选取这些链路进行迁移处理。

（2）资源评估与路径搜索：针对每条选定的故障虚拟链路，筛查物理链路资源，剔除那些无法满足其带宽需求的物理路径^[10]。随后，运用最短路径算法，在剩余的物理网络中探寻一条直接连接该虚拟链路两端节点的物理路径。若成功找到，则执行迁移操作；若未果，则转入路径分割迁移流程。

（3）路径分割与迁移：对于需要进行分割迁移的故障虚拟链路，再次运用最短路径算法，全面搜索并列出现有连接其两端物理节点的独立物理路径集合 L 。随后，根据这些路径中剩余的最小可用带宽进行从大到小的排序。

通过模拟不同分割方案下的网络性能，选择最优路径组合进行迁移，确保故障虚拟链路在恢复过程中既能保持高效传输，又能减少因单点故障再次影响整体网络稳定性的可能性。此外，算法还集成了实时监控与动态调整机制，根据网络状态变化灵活调整迁移策略，确保高精度虚拟重构的连续性和有效性。

5 测试实验

5.1 实验准备

在图 1 所示的物联网高精度虚拟重构算法实验准备阶段，设计了节点与缺失信息的映射。节点映射方案设定为 $\{X \rightarrow \alpha, Y \rightarrow \beta, Z \rightarrow \gamma\}$ ，旨在将实际物联网中的缺失信息采集点映射至虚拟重构空间中的对应位置。同时，规划了链路映射方案，具体为 $\{(X, Y) \rightarrow (\alpha, \beta), (Y, Z) \rightarrow (\beta, \gamma)\}$ ，确保缺失信息间的潜在关联在虚拟环境中得以准确重建重构。

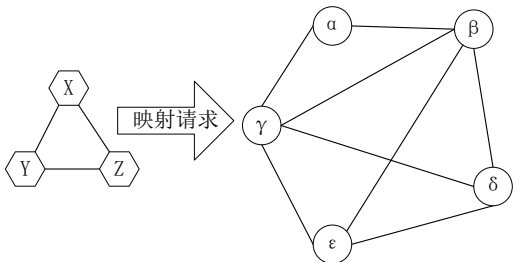


图 1 节点与缺失信息的映射

根据上述实验准备，为了验证高精度虚拟重构算法在物联网环境下处理采集数据缺失问题的效能与实用性，构建了一个高性能且配置精良的实验平台，参数如表 1 所示。

表 1 具体参数

序号	参数类别	具体参数
1	处理器型号	AMD Ryzen 5
2	图像捕获单元配置	双核 800 MHz
3	图像处理器	主频 2.5 GHz 内存 4.00 GB
4	操作系统	Windows 10
5	编程语言	Python
6	图像处理库	OpenCV
7	D/A 转换器采样率	4.2 Hz
8	光圈设置	F16
9	曝光时间	15 s
10	A/D 转换范围	-2~6 dB

具体的实验现场如图 2 所示。



图 2 实验环境

为了验证本文提出的重构算法在物联网环境下处理缺失信息的有效性和高精度，设计并实施了一系列详尽的仿真实

验。实验中，特别配置了 Visual DSP 环境，以模拟真实的物联网数据传输和处理场景。依据精心预设的仿真参数，成功构建了包括随机丢失、周期性缺失以及突发故障导致的数据中断在内的多种数据缺失场景，这些场景紧密贴合物联网实际应用中可能遇到的各种复杂情况。采用设计的高精度虚拟重构算法，并特别配置了 Visual DSP 环境以进行详尽的仿真测试。依据实验预设的仿真参数，成功模拟了包括随机丢失、周期性缺失以及突发故障导致的数据中断在内的多种数据缺失场景，这些场景紧密贴合实际物联网应用中可能遇到的问题。在仿真测试过程中，本文提出的重构算法展现出了卓越的虚拟重构性能能力，该方法能够迅速响应并识别出原始的缺失信息样本，它能够迅速识别并分析初始缺失的信息样本，如表 2 所示。

表 2 物联网缺失信息样本

传感器编号	数据类型	预期值	实际接收值	是否缺失
S1	温度 /℃	23.5	23.5	否
S2	湿度 /%	65	—	是
S3	光照强度 /Lux	500	500	否
S4	气压 /hPa	1 013.25	1 013.25	否
S5	气体浓度 /(mg·m ⁻³)	50	50	否
S6	振动强度 /Hz	3.2	—	是
S7	电流 /A	0.8	0.8	否
S8	电压 /V	24	—	是

为了构建一个多元化的模拟数据集，以贴近实际物联网环境中复杂多变的数据缺失情况，构建一个多元化的模拟数据集，在数据集的构建过程中，保留了表 2 中的基础信息作为核心框架。在此基础上，引入随机生成机制，以模拟实际物联网环境中可能发生的随机性数据缺失现象，这些缺失值在时间序列上呈现出无规律分布。通过调整缺失比例，设置了从轻微的数据遗漏到严重的数据断层等多个不同的缺失层级。每一层级都代表着物联网系统在特定运行状态下可能遭遇的不同程度的数据缺失状况。对于每一个缺失层级，都采用了不同的随机算法和参数配置，以生成具有特定缺失特性的数据子集。这些子集不仅保留了表 2 中的关键信息字段，还在其时间序列数据中随机插入了缺失值，且缺失值的分布和频率均根据缺失层级的设定进行精确控制。通过这一流程，生成的多个扩展数据集，如表 3 所示。

表 3 扩展数据集

编号	缺失比例	缺失类型	传感器编号	数据类型
DS1	10%	随机	S1-S8	所有类型
DS2	20%	连续	S2、S4、S6	湿度、气压、振动
DS3	5%	块状	S1、S3、S5	温度、光照、气体
DS4	30%	随机 + 连续	S1-S8	所有类型
DS5	15%	块状 + 随机	S2、S5、S7	湿度、气体、电流

5.2 实验结果分析

为了全面评估不同算法在物联网数据缺失场景下的性能表现,选取了三种具有代表性的算法进行对比分析,分别为文献[1]提出的基于深度学习与压缩感知理论的通用图像重构算法、文献[2]提出的偏振光谱多维信息的重构融合算法,以及本文创新提出的算法。为了确保实验结果的稳定性,减少数据随机性对评估结果的影响,对每种算法进行计划进行30次独立仿真实验。在每次实验中,均模拟物联网环境中的数据缺失情况,并应用相应的算法进行重构,并计算这30次实验结果的均值,以此作为评估各算法性能的最终评估指标。3种不同算法在处理物联网采集缺失信息时的请求接受率变化情况,如图3所示。

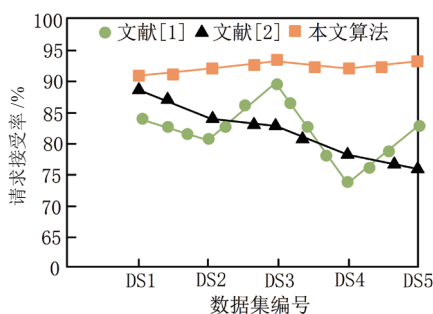


图3 物联网采集缺失信息请求接受率

5.3 实验结果分析

根据图3可知,所有算法在处理这五个不同数据集(DS1至DS5)时均表现出一定的有效性,但它们在性能上存在显著差异。本文提出的算法在所有五个数据集上的请求接受率都均显著高于文献[1]和文献[2]中的算法,特别是在DS3数据集上,其请求接受率高达93.2%,这一数值显著优于其他两种算法,充分证明了本文算法在复杂物联网环境下具有出色的高精度重构能力。这一优势体现了算法对不同数据特性和缺失模式的高度适应性。相比之下,文献[1]算法虽然也展现了不错的性能,但在处理DS4数据集时,其请求接受率相对较低,这可能是因为该算法对特定数据类型和缺失模式的适应性有限。而文献[2]算法,虽然在DS2数据集上表现良好,但从整体来看,其在稳定性和平均性能上略逊于本文算法。

综上所述,随着数据集多样性的不断增加,本文算法的优势越发显著。它不仅能够有效应对复杂多变的物联网数据,还在实际应用中展现出了强大的适应性和稳定性。

6 结语

本研究针对物联网数据采集中普遍存在的信息缺失问题,提出了一种创新的高精度虚拟重构算法,该算法不仅成功融合了MapReduce框架的并行处理能力,实现了数据的高效预处理与特征提取,还通过设计先进的缺失信息特征识别

器,显著提升了缺失数据的识别精度。在构建虚拟重构模型时,算法充分考虑了实际物联网环境中的多重约束条件,并借助历史数据的深度训练,展现出了对复杂数据缺失情况的强大适应能力。实验结果表明,对不同数据特性和缺失模式展现出了高度的适应性,尤其是DS3数据集上高达93.2%的请求接受率,充分验证了该算法在复杂物联网环境下的高精度重构能力。因此,本研究不仅为物联网数据处理领域贡献了一种高效、可靠的解决方案,也为未来物联网技术的进一步发展提供了新的思路与方向。在物联网时代,面对数据采集的种种挑战,本文提出的高精度虚拟重构算法展现了强大的数据处理与恢复能力,有效解决了物联网环境中信息缺失的难题,为数据的完整性提供了坚实保障。展望未来,随着技术的不断进步和应用场景的持续拓展,这一算法将在更多领域发挥重要作用,推动物联网技术迈向新的高度。

参考文献:

- [1] 郭媛,姜津霖.基于深度学习与压缩感知理论的通用图像重构算法[J].黑龙江大学学报,2024,15(2):34-41+2.
- [2] 钟菁菁,刘骁,王雪霁,等.偏振光谱多维信息的重构融合算法[J].光谱学与光谱分析,2023,43(4):1254-1261.
- [3] 周晓艳,唐涛,张思乾,等.多角度SAR图像非目标遮挡缺失信息重构[J].信号处理,2021,37(9):1569-1580.
- [4] 龙雄辉,苏丹.城市轨道交通车辆管控虚拟重构方法设计[J].传感器世界,2024,30(3):14-18+26.
- [5] 赵太飞,张健伟,容开新,等.紫外非均匀虚拟势场下无人机编队重构避碰算法[J].电子与信息学报,2023,45(5):1651-1659.
- [6] 胡青松,王胜男.一种煤矿井下无线自组网灾后重构算法[J].工矿自动化,2022,48(2):93-99.
- [7] 张磊,缪永伟,景程宇,等.融合缺失点云形状信息的保结构修复网络[J].计算机辅助设计与图形学学报,2023,35(5):696-707.
- [8] 吴晗,王士同.不完整数据分类与缺失信息重要性识别特权LSSVM[J].智能系统学报,2023,18(4):743-753.
- [9] 程朋欢.数据拟合与数据插值在缺失信息补充中的精度分析[J].测绘与空间地理信息,2022,45(12):127-129.
- [10] 谢宇,黄鹏,李震,等.改进未知互耦信息下矩阵重构算法的波束形成[J].无线电通信技术,2022,48(6):1092-1097.

【作者简介】

米婉晋(1995—),女,山西晋中人,硕士,助教,研究方向:物联网、虚拟现实。

(收稿日期:2024-08-01)