基于属性基加密的物联网隐私数据安全共享研究

刘晓蒙[1]
LIU Xiaomeng

摘 要 在物联网领域，设备数量庞大且持续产生海量、多样化的数据，这类数据的复杂性使得直接对其进行共享与处理面临巨大挑战。为有效解决这一问题，文章提出一种基于属性基加密技术的物联网隐私数据安全共享算法。该算法首先应用模糊 C 均值聚类算法，通过迭代更新聚类中心直至目标函数收敛，实现物联网隐私数据的聚类处理，以此应对设备繁多、数据海量且多样化的问题；接着通过哈希函数与数字签名技术，对聚类后的物联网隐私数据进行完整性验证，防止数据在传输或存储过程中被未经授权篡改或破坏，确保接收方获取数据的真实性与可靠性；最后引入属性基加密算法实现验证后数据的安全共享，生成公共参数与主密钥，数据发送者利用公共参数和访问策略对数据加密，用户则通过自身私钥与公共参数对密文解密，通过完整的加解密过程保障数据共享安全。实验结果表明，所提算法能够有效聚类不同类型数据，不仅吞吐量较高，且可确保仅满足访问策略的用户能解密数据，从而保障物联网隐私数据在共享中的安全性，整体性能较为优异。

关键词 属性基加密；物联网；隐私数据；安全共享；模糊 C 均值聚类算法

doi：10.3969/j.issn.1672-9528.2025.09.033

0 引言

随着物联网（IoT）技术的快速发展，智能设备数量呈激增态势，大量隐私数据（如个人健康信息、位置数据、消费习惯等）随之被采集、传输与存储。这些数据在提升生活便利性和智能化水平的同时，也面临着严重的隐私泄漏风险[1]。传统的数据共享方法通常依赖于集中式存储和加密技术，存在单点故障、数据滥用和权限管理不足等问题，难以满足物联网环境下大规模、动态性和多样性的隐私保护需求[2]。因此，研究物联网隐私数据安全共享方法具有重要的现实意义，能够在确保数据隐私的前提下，实现数据的高效共享与利用，推动物联网技术在智慧城市、智能医疗等领域的健康发展，同时为构建安全可信的物联网生态系统提供技术支撑。

近几年，相关研究者针对物联网隐私数据安全共享[3-4]方面的内容展开了大量研究，例如李井涵等人[5]将全部信息存储到区块中，有效防止恶意用户未经授权访问设备，实现数据安全共享，但是区块链的透明性和公开性在某些情况下会与隐私保护需求相冲突，不能确保隐私数据的安全性充分保证。潘雪等人[6]同样通过区块链技术实现物联网隐私数据安全共享，但是在面对规模比较大的数据时，区块链会出现运行缓慢的情况，造成吞吐量较低。Wang 等人[7]通过智能椭圆数字签名算法完成隐私信息安全共享，该方法在应用过程中需要重点考虑将密钥分发给合法的用户或设备，如果分配不当将会出现密钥泄漏或被非法获取的情况，从而影响隐私数据的安全性。傅文龙等人[8]主要通过国产加密算法完成密钥交换，然后引入智能合约实现数据和主题的共享以及管理。智能合约的编写需要高度精确，任何错误都可能导致数据泄漏，影响数据安全共享性能。

鉴于当前隐私数据安全共享研究中存在的挑战，本研究旨在探索并设计一种基于属性基加密的物联网隐私数据安全共享算法，以更高效、灵活地实现数据的安全共享与访问控制，同时保障数据的隐私性。

1 一种物联网隐私数据安全共享算法设计

1.1 物联网隐私数据动态聚类

由于物联网设备众多，产生的数据量庞大且复杂多样，直接对这些数据进行共享处理可能面临巨大的挑战。因此，需要对物联网隐私数据进行聚类。通过聚类分析，可以将相似的数据进行合并处理，减少数据的冗余和重复，提高数据共享的效率。

从物联网存储系统中提取隐私数据，并定义隐私数据二元语义特征向量为$\boldsymbol{v}=\left[v_{i1},v_{i2},...,v_{in}\right]$，引入特征空间重组技术重构隐私数据，精准抽取关联信息特征量，计算出隐私数据

1. 郑州升达经贸管理学院信息工程学院 河南郑州 451191

的聚类加权系数：

$$w_i = g(\boldsymbol{v}_i') \tag{1}$$

式中：w_i 代表第 i 个数据的聚类加权系数；g 代表加权系数计算函数；$\boldsymbol{v}_i'$ 代表重构后的特征向量。

基于聚类加权系数识别物联网隐私数据的聚类中心，可以表示为：

$$c_k = \frac{\sum_{i \in C_k} w_i \cdot \boldsymbol{v}_i'}{\sum_{i \in C_k} w_i} \tag{2}$$

式中：c_k 代表第 k 个聚类中心；C_k 代表属于第 k 个聚类的数据集合。

完成上述操作后，采用模糊 C 均值聚类算法通过迭代更新聚类中心，直到目标函数收敛或达到预定的迭代次数，模糊 C 均值聚类的目标是最小化以下目标函数：

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^m \left\| x_i - c_j \right\|^2 \tag{3}$$

式中：N 代表数据点的总数；C 代表聚类的数量；u_{ij}^m 代表第 i 个数据点对第 j 个聚类的隶属度；m 代表模糊因子；x_i 代表第 i 个数据点；c_j 代表第 j 个聚类中心。

通过式（4）对数据聚类中心进行更新：

$$c_k = \frac{\sum_{i=1}^{N} u_{ij}^m x_i}{\sum_{i=1}^{N} u_{ij}^m} \tag{4}$$

通过迭代更新聚类中心，使目标函数收敛或达到预定的迭代次数，由此实现物联网隐私数据聚类。

考虑到模糊 C 均值聚类算法需要一次性处理所有数据，无法适应数据流环境。因此，设计一种能够动态更新聚类中心的机制，以实时处理新数据并调整聚类结果。该机制的核心思想是：在数据流环境中，每当新数据到达时，动态更新聚类中心，而无需重新计算所有数据的聚类结果。

当新数据点 x_{n+1} 到 +1 达时，计算其对每个聚类中心 c_j 的隶属度 $x_{(n+1)j}$：

$$x_{(n+1)j} = \frac{1}{\sum_{l=1}^{k} \left(\frac{\left\| x_{n+1} - c_j \right\|}{\left\| x_{n+1} - c_l \right\|} \right)^{\frac{2}{m-1}}} \tag{5}$$

根据新数据点的隶属度，动态更新聚类中心 c_j：

$$c_j = \frac{\sum_{i=1}^{n+1} u_{ij}^m x_i}{\sum_{i=1}^{n+1} u_{ij}^m} \tag{6}$$

式中：u_{ij} 为数据点 x_i 对聚类中心 c_j 的隶属度。

重复上述步骤，直到聚类中心的变化小于预设阈值 ε，即 $\left\| c_j^{\text{new}} - c_j^{\text{old}} \right\| < \varepsilon$。通过引入在线聚类机制，可以有效解决物联网数据动态性和实时性带来的挑战，为物联网隐私数据安全共享提供更高效、更灵活的解决方案。

1.2 物联网隐私数据完整性验证

模糊 C 均值聚类算法将相似数据分组后，为了防止数据在传输或存储过程中被未经授权的篡改或破坏，确保接收方获得的数据是真实可靠的。同时，为了防止敏感信息泄漏给未经授权的第三方，维护用户的合法权益和信息安全，从而保障数据共享的安全性，通过哈希函数和数字签名技术进行物联网隐私数据完整性验证。

将聚类后的数据集分成不同的簇，对每个簇或簇内的关键数据应用哈希函数，生成哈希值用于验证数据的完整性。假设哈希函数为：

$$H : \{0,1\}^* \to \{0,1\}^l \tag{7}$$

式中：$\{0,1\}^*$ 表示任意长度的二进制字符串；$\{0,1\}^l$ 表示长度为 l 的二进制字符串。

对于任意簇中的数据点 x，计算其哈希值 $h_x = H(x)$。验证时，接收方重新计算数据的哈希值并与发送的哈希值比较，若相等则认为数据未被篡改。

在数据传输前，使用数字签名技术对哈希值进行签名，以确保数据来源的真实性和完整性。

假设私钥为 sk，公钥为 pk，对哈希值 h_x 进行签名，生成签名：

$$\sigma = \text{Sign}_{\text{sk}}(h_x) \tag{8}$$

验证签名时，使用公钥 pk 和签名 σ，验证 $\text{Verify}_{\text{pk}}(\sigma, h_x)$ 是否为真。

通过聚类处理将相似数据分组，减少了后续哈希处理的计算量，并提高了哈希值的区分度（因为相似数据被分组处理）。哈希函数的应用确保了数据的完整性，即使数据被篡改，哈希值也会变化，从而被检测出来。数字签名则是进一步加密认证，确保了数据的来源真实性和完整性。没有正确的私钥，无法伪造有效的签名，因此接收方可以通过验证签名来确认数据的真实性和完整性。

综上所述，通过聚类、哈希函数和数字签名技术的结合，可以有效地保护物联网隐私数据在传输过程中的完整性和隐私性。

1.3 物联网隐私数据安全共享

完成物联网隐私数据完整性验证 [9-10] 处理后，引入属性基加密算法对验证后的物联网隐私数据展开安全共享研究，属性基加密（attribute-based encryption, ABE）是一种加密技术，它允许数据根据用户的属性进行加密和解密。在物联网隐私数据安全共享的场景中，属性基加密可以确保只有具备特定

属性的用户才能解密数据。以下是物联网隐私数据属性基加密的具体过程：

（1）系统初始化阶段生成公共参数 P 和主密钥 MK，公共参数对所有用户公开，而主密钥由可信的密钥生成中心（key generation center, KGC）保存，主密钥用公式表示为：

$$\mathrm{MK}=\left(g^{a},\left\{g^{a_i}\right\}_{i=1}^{n}\right) \tag{9}$$

式中：a 是一个随机数；a_i 是与第 i 个属性相关的随机数。

（2）数据发送者使用公共参数和访问策略对数据进行加密：

$$\mathrm{CT}=\left(\mathrm{Enc}(P,\mathrm{AS}),\mathrm{AS}\right) \tag{10}$$

式中：Enc 是加密算法；AS 是访问策略；CT 是密文。

（3）用户向 KGC 提供自己的属性集，KGC 根据属性集生成用户的私钥：

$$\mathrm{SK}=\left(\left\{D_i=g^{\frac{a_i}{b_i}}\right\}i\in S,D=g^{\frac{1}{b}}\right) \tag{11}$$

式中：S 是用户的属性集；b 是一个随机数；b_i 是与属性 i 相关的随机数。

（4）用户使用自己的私钥和公共参数对密文进行解密：

$$M=\mathrm{Dec}(P,\mathrm{SK},\mathrm{CT}) \tag{12}$$

式中：Dec 是解密算法。

属性基加密确保了只有满足访问策略的用户才能解密数据，从而实现了物联网隐私数据的安全共享。

在物联网环境中，不同用户可能具有不同的权限级别。例如，管理员可以访问所有数据，普通用户只能访问部分数据，而访客只能访问公开数据。上述 ABE 算法通常只支持单一层次的访问控制，无法满足这种多级权限的需求。因此，设计一种多级访问控制机制，确保高权限用户能够访问更多数据，而低权限用户只能访问有限数据。

（1）假设系统中有 N 个权限级别，分别为 $L_1, L_2,\cdots, L_N$，其中，L_1 为最高权限级别；L_N 为最低权限级别。通过权限级别划分，实现更精细的数据访问控制，满足不同用户的需求。

（2）根据用户的权限级别，动态调整其能够访问的数据范围。高权限用户可以访问更多或更敏感的数据，而低权限用户只能访问有限或公开的数据。根据用户的权限级别动态调整访问策略，适应复杂多变的实际应用场景。

通过引入多级访问控制机制，ABE 算法能够更好地适应复杂应用场景中的权限管理需求。

2 实验分析

为了验证所提基于属性基加密的物联网隐私数据安全共享算法的有效性，选取区块链法和主从链法作为对比算法展开对比实验分析。实验中，迭代次数设置为 1 000 次，物联网的覆盖范围为 500 m×500 m。

实验数据来源于 UCI Machine Learning Repository，样本规模为 1 200 个，数据属性为 10 个，涵盖了与物联网隐私直接相关的数据类型，以确保实验数据的针对性和实用性。1 200 个样本中训练集占 80%（960 个样本），测试集占 20%（240 个样本）。数据属性为 10 个，包括 5 个数值型属性（用户 ID、设备 ID、时间戳、地理位置坐标、数据访问频率）和 5 个类别型属性（用户角色、设备类型、数据敏感级别、访问权限、数据来源）。类别型属性中，用户角色包括管理员、普通用户、访客，设备类型包括智能家居、工业设备、医疗设备，数据敏感级别包括低、中、高。

为了验证所提算法在物联网隐私数据聚类方面的优势，选取 K-modes 聚类算法和有向图聚类算法作为对比算法，其中，K-modes 聚类算法是 K-means 算法的变体，它通过计算汉明距离来度量数据点之间的不相似性，以此实现数据聚类。有向图聚类算法则是基于图论的方法，它将数据点视为图中的节点，并通过构建有向边来表示数据点之间的关系，然后利用图的结构特性来进行聚类。通过与上述两种方法进行对比可以更全面地评估所提算法在物联网隐私数据聚类中的性能和适用性。结果如图 1 所示。

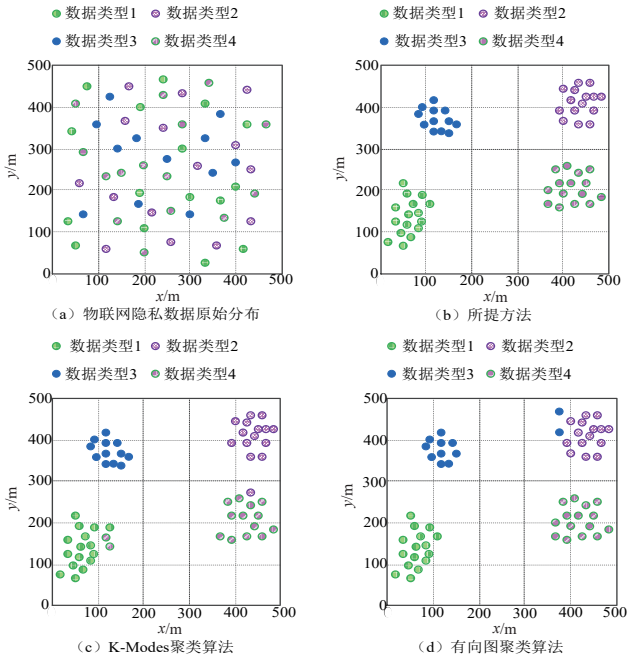


图 1 不同算法的物联网隐私数据聚类效果比较

分析图 1 可以看出，在相同的实验场景下，采用所提算法可以精准对各个类型的物联网隐私数据展开有效聚类，不存在数据混淆的问题，而 K-modes 聚类算法和有向图聚类算法在聚类过程中出现了错误聚类的情况，整体的聚类性能不

佳，不利于后续的物联网隐私数据安全共享。由于引入了模糊性，模糊 C 均值聚类算法对初始聚类中心的选择相对不敏感。这意味着算法在不同初始条件下通常能够收敛到相似的聚类结果，减少了因初始条件选择不当而导致的聚类误差，因此，数据聚类效果较好。

在物联网环境中，由于设备数量庞大且数据产生速度快，高吞吐量意味着算法能够快速响应并处理大量的隐私数据，从而提高数据处理的效率和实时性。因此，在相同实验环境下，分析区块链算法、主从链算法和所提算法的吞吐量变化情况，结果如图 2 所示。

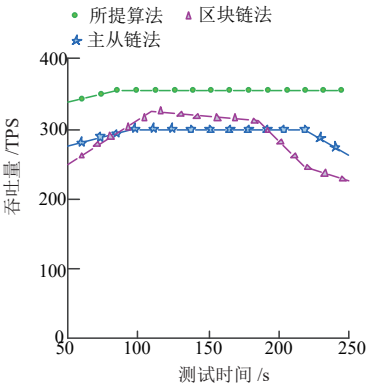
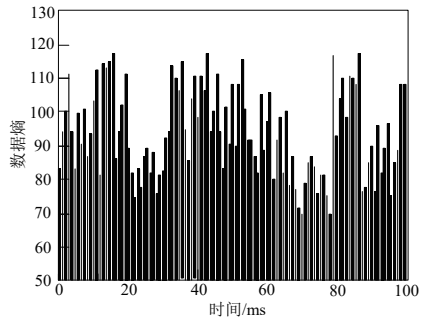


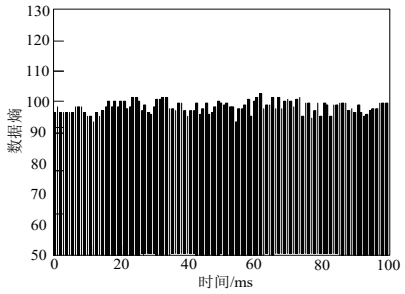
图 2 不同算法的吞吐量测试结果比较

分析图 2 可以看出，所提算法的吞吐量在初始阶段呈现缓慢上升的趋势，然后逐渐趋于稳定，可高达 340 TPS 以上；另两种算法在吞吐量表现上起初展现出缓慢增长的态势，随后趋于稳定状态，但最终呈现出下降的趋势，且明显低于 340 TPS。由此可见，所提算法能够快速响应并处理大量的隐私数据，说明其性能最优。这是因为所提算法通过模糊 C 均值聚类算法将物联网隐私数据进行聚类，从而有效减少了后续数据处理和加密的复杂度，显著提高了系统的吞吐量，使得物联网平台在面对设备繁多、数据量大且类型多样的挑战时，能够更快速、更顺畅地处理并共享隐私数据，满足实时性和高效性的需求。

为了进一步验证所提算法在物联网隐私数据安全共享方面的优越性，给出其数据加密结果，如图 3 所示。



（a）加密前



（b）加密后

图 3 所提算法物联网隐私数据安全共享性能测试结果

分析图 3 可以看出，未经加密处理的数据呈现出多样化的数据熵分布特征，这一特性使得根据数据的独特属性精准定位并提取目标数据成为可能。经过所提算法加密后，其原有的熵分布模式发生了显著变化，趋向于统一且均衡的状态，极大地缩小了数据项之间的差异性。这种变化模糊了数据间的界限，从而构建了一道坚实的安全屏障。具体而言，所提算法确保了智能电网半结构化数据共享的高度安全性，有效抵御了潜在的数据泄漏风险。

非法访问检测率直接反映了数据安全共享方法在保护数据隐私和防止数据泄漏方面的有效性，区块链算法、主从链算法和所提算法的非法访问检测率对比结果如图 4 所示。

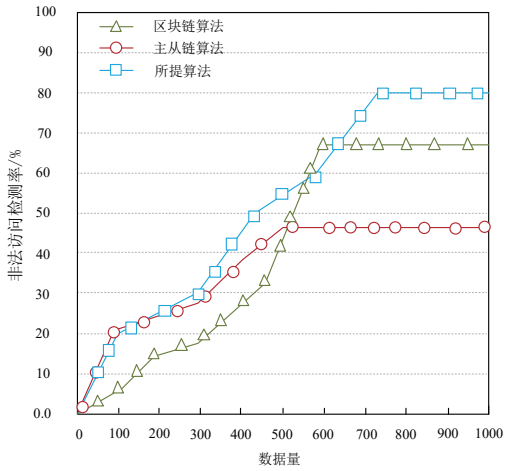


图 4 非法访问检测率对比结果

从图 4 的对比结果来看，随着数据量的增加，区块链算法、主从链算法和所提算法的非法访问检测率均呈现上升趋势。所提算法在各数据量下的非法访问检测率显著高于区块链算法和主从链算法。当数据量为 100 时，所提算法的检测率就已经领先；随着数据量逐步增大到 1 000，所提算法的检测率优势愈发明显，持续保持在高位，而区块链算法和主从链算法的检测率虽然也在增长，但始终低于所提算法。这表明所提算法在识别和阻止非法访问方面更为有效，在保护物联网隐私数据安全共享上具有更强的能力和优势，相比区块链算

法和主从链算法更能满足对数据隐私保护和防止数据泄漏的要求。

隐私泄漏概率是指在攻击条件下，隐私数据被泄漏的概率。通过模拟攻击实验，统计成功泄漏隐私数据的次数与总攻击次数的比值。区块链算法、主从链算法和所提算法的隐私泄漏概率对比结果如图 5 所示。

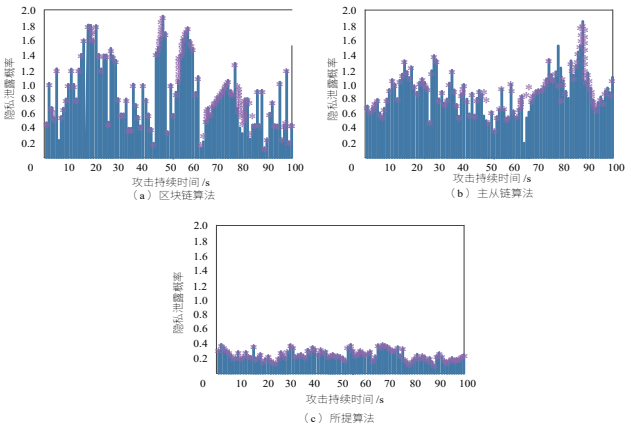


图 5 隐私泄漏概率对比结果

从图 5 的隐私泄漏概率对比结果来看，在模拟攻击实验中，区块链算法、主从链算法和所提算法呈现出明显不同的表现。区块链算法和主从链算法的隐私泄漏概率相对较高，且波动幅度较大，在某些攻击时刻，其隐私泄漏概率数值达到较高水平，这意味着使用这两种算法时，隐私数据面临着较大的泄漏风险。相比之下，所提算法的隐私泄漏概率则始终维持在极低的水平，几乎趋近于零，且波动幅度极小。这充分表明所提算法在抵御攻击、保护隐私数据方面具有显著优势，能够更有效地降低隐私数据被泄漏的可能性，在物联网隐私数据安全共享场景中，其安全性表现远超区块链算法和主从链算法。

综上所述，通过应用所提算法智能电网半结构化数据的安全性得到了显著提升，不仅保护了数据的机密性，还降低了数据被非法利用的风险，为智能电网的可靠运行和电力行业的数字化转型提供了强有力的支持。

3 结语

物联网技术的飞跃进步正引领着全面互联时代的到来，海量数据的产生与交换成为常态，但同时也伴随着前所未有的隐私泄漏风险。为此，提出一种基于属性基加密的物联网隐私数据安全共享算法：

（1）将模糊 C 均值聚类算法应用于物联网隐私数据的预处理阶段，通过聚类分析有效应对了设备繁多、数据海量且多样化的挑战，提高了数据处理和共享的效率。

（2）引入哈希函数和数字签名技术，在数据聚类后、加密前进行完整性验证，确保数据在传输或存储过程中未被篡改或破坏。这一机制增强了数据共享的可靠性，为接收方提供了数据真实性的保障。

（3）引入属性基加密技术，为聚类和验证后的数据提供细粒度的访问控制，实现了数据的安全共享。这种结合聚类、验证与加密的方法，既减少了加密的数据量，又提高了访问控制的灵活性。

（4）通过大量实验分析证明，所提算法可以显著提升物联网隐私数据聚类效果，提升数据处理效率，同时还可以有效确保物联网隐私数据的安全性和可靠性，更好地实现物联网隐私数据安全共享。

参考文献：

[1] 梁盈威，杨秋勇，谢瀚阳 . 物联网环境下数据开放性共享安全保障体系 [J]. 微型电脑应用 ,2022,38(9):194-197.

[2] 朱雪岭，侯慧莹，付绍静，等 . 面向便携式诊所的安全数据共享方案 [J]. 软件学报 ,2023,34(9):4256-4274.

[3] 冯涛，陈李秋，方君丽，等 . 基于本地化差分隐私和属性基可搜索加密的区块链数据共享方案 [J]. 通信学报 , 2023, 44(5): 224-233.

[4] 牛淑芬，宋蜜，方丽芝，等 . 智慧医疗中基于属性加密的云存储数据共享 [J]. 电子与信息学报 ,2022,44(1):107-117.

[5] 李井涵，沈国华，杨阳，等 . 结合区块链的物联网数据安全共享机制 [J]. 小型微型计算机系统 ,2023,44(8):1812-1818.

[6] 潘雪，袁凌云，黄敏敏 . 主从链下的物联网隐私数据跨域安全共享模型 [J]. 计算机应用研究 ,2022,39(11):3238-3243.

[7] WANG Y, CHE T Y, ZHAO X H, et al. A blockchain-based privacy information security sharing scheme in industrial internet of things[J]. Sensors, 2022, 22(9): 3426.

[8] 傅文龙，李国刚，解童 . 采用区块链的物联网数据共享方案 [J]. 华侨大学学报 (自然科学版),2023,44(2):257-263.

[9] 张国鹏，陈学斌，王豪石，等 . 面向本地差分隐私的 K-Prototypes 聚类方法 [J]. 计算机应用 ,2022,42(12):3813-3821.

[10] 张东月，倪巍伟，张森，等 . 一种基于本地化差分隐私的网格聚类方法 [J]. 计算机学报 ,2023,46(2):422-435.

【作者简介】

刘晓蒙（1993—），男，河南郑州人，硕士，讲师，研究方向：数据挖掘算法原理。

（收稿日期：2025-03-31 修回日期：2025-09-04）