人工智能与产业融合

基于迭代模糊聚类的自然资源现状数据集成

王兴娟，姬 祥

（济宁市国土空间数据和遥感技术中心，济宁 272000）

摘要：为有效解决自然资源现状数据集成过程中面临的类别混淆、规范性不足及处理效率低下等关键问题，提出基于迭代模糊聚类的自然资源现状数据集成方法。通过标准指标法与“内梅罗”综合指标法构建“标准指数-综合指数”分级体系，对全域全要素自然资源现状数据进行科学分类；进而引入迭代模糊聚类算法对已分类数据进行集中处理，通过优化模糊权值与迭代过程，在显著减少训练样本量的同时，提升了处理器训练效率与聚类准确率。在此基础上，利用处理后的数据构建全域自然资源信息矩阵，并设计基于数据相关性的聚类处理流程，实现同类属性数据的高精度集成与异常数据的有效筛选。实验结果表明，该方法在数据集成归一化处理方面表现优异，处理后的数据规律性显著、同类别间相似度高，有效满足了自然资源现状分析的实践需求，为构建精准高效的自然资源业务全要素信息化体系提供了方法论支撑。

关键词：迭代模糊聚类；集成方法；现状数据；自然资源；双层级分类体系

中图分类号：TP242 **文献标志码**：A **文章编号**：1671－1807(2026)16－0024－07

自然资源的科学开发与可持续利用，是应对全球人口增长、资源短缺及环境污染等严峻挑战的关键途径[1]。全面、精准地掌握自然资源现状，是实现科学决策与高效管理的重要基础。在此背景下，全域、全要素、多源异构的自然资源现状数据的高质量集成，已成为支撑自然资源“一张图”管理与业务全要素信息化的核心前提[2]。

目前，针对多源异构数据集成，已有诸多研究尝试。孙中彬等[3]提出基于安全欠采样的不均衡多标签数据集成方法，通过标记分割与集成学习处理数据不平衡问题，但该方法在处理海量数据时，算法复杂度呈指数级增长，计算复杂度高，对计算资源消耗极大，且易丢失关键信息，难以适用于海量自然资源数据场景。王江江等[4]在能源系统优化中开展了多能互补的集成研究，其模型对数据稳定性要求较高，在处理自然资源监测数据时，为保证数据稳定性进行大量预处理工作，增加了资源消耗，且参数设定对结果影响显著，在波动性强的自然资源监测数据中适应性不足。由此可见，面对自然资源数据来源多样、结构复杂、尺度不一的特点，现有方法仍普遍存在类别混淆率高、集成规范性不足、处理效率低下等问题。

为有效解决现有方法在类别区分精度与集成效率方面的不足，本文引入迭代模糊聚类算法，旨在通过构建“标准指数-综合指数”分类体系实现对数据的科学归类，并结合数据相关性聚类处理以构建全域自然资源信息矩阵，从而显著提升数据集成的内在一致性、类别纯净度与处理效能，为构建精准、高效的自然资源业务信息化体系提供新的方法支撑。

1 数据和方法

1.1 多类型自然资源现状空间数据集类别划分

为满足自然资源现状数据集成需求，本文设计方法收集多类型自然资源现状空间数据，这些数据涵盖土地利用类型、植被覆盖度、土壤类型与质地、地形地貌特征、气候数据、水文数据、生物多样性数据及矿产资源分布等多个方面[5]。为了实现这些数据的科学分类，引进标准指标法进行数据中单一因素的评估。具体而言，以多类型自然资源现状空间数据集中的资源类型、资源类别、属性数量、生物体等为研究对象，通过对比各个因素的现状数据结果，得出对应数据的相关基准指标，并以此为基础确定自然资源的标准指数，具体计算公式如下[6]。

收稿日期：2026-03-31

作者简介：王兴娟(1986—)，女，山东临沂人，工程师，研究方向为自然资源信息化、自然资源数据治理、新型智慧城市建设；通信作者姬祥(1985—)，男，山东济宁人，高级工程师，研究方向为自然资源信息化、自然资源数据治理、网络安全管理等。

$$K_0=\frac{M_1}{M_2} \tag{1}$$

式中:K_0 为标准指数;M_1 为全域全要素自然资源的指标实测结果;M_2 为全域全要素自然资源的指标监测结果。

根据自然资源的种类,结合上文计算得到的标准指数,采用“内梅罗”指标法,计算自然资源数据的综合指标[7],计算方法如式(2)所示。

$$K_k=\sqrt{\frac{\max(K_{0i})^2+(\mathrm{ave}K_{0i})^2}{2}} \tag{2}$$

式中:K_k 为自然资源数据的综合指标;$K_{0,i}$ 为第 i 种资源的标准指数;ave 为平均值计算。以标准指数为参照进行数据的筛选,将综合指数作为划分的依据,实现自然资源现状数据类别的划分。根据自然资源的特性和实际需求,通过“标准指数-综合指数”分级体系,将全域全要素的自然资源现状数据划分为多个具有明确意义的类别,这些类别直接对应不同的自然资源类型和属性,为后续的数据集成和处理提供了清晰的框架[8]。

在本文中,实验数据范围广泛,覆盖全域全要素的自然资源现状数据,具体包括但不限于土地利用类型、植被覆盖度、土壤类型与质地、地形地貌特征、气候数据、水文数据、生物多样性数据及矿产资源分布等。这些数据通过“浮标定点监测-船载快速监测-航空遥感-卫星遥感”相结合的技术手段进行采集,确保了数据的全面性和准确性[9]。例如,土地利用类型数据详细记录了各类用地的空间分布与面积比例;植被覆盖度数据通过高分辨率遥感影像解译获得,反映了地表植被的茂密程度;气候与水文数据通过长期观测站网络持续收集,为分析自然资源时空变化提供了重要依据。

1.2 基于迭代模糊聚类的数据处理

在上述基于“标准指数-综合指数”完成数据类别初步划分的基础上,引入迭代模糊聚类算法对各类别内的数据进行精细化处理。迭代模糊聚类本质上是一种软聚类方法。与传统硬聚类不同,迭代模糊聚类允许一个数据点以不同的隶属度同时隶属于多个簇,更适用于自然资源数据中普遍存在的边界模糊、属性重叠的实际情况[10]。其核心思想是通过迭代优化,动态寻找能够最佳表征数据分布的簇中心及每个数据点对各个簇的隶属度。数据处理流程如图 1 所示。

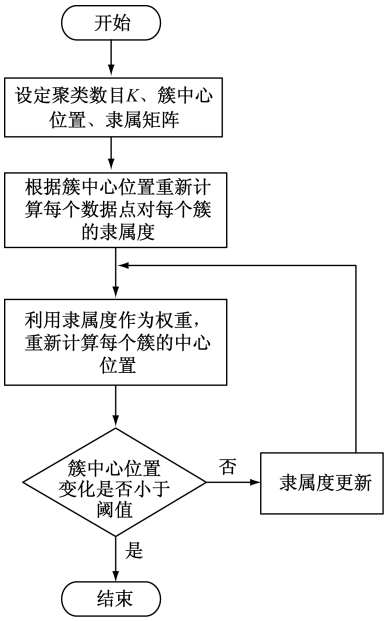


图 1 数据处理流程

基本的迭代优化过程包括以下几个关键步骤。

步骤 1:初始化。确定或随机设定期望的聚类数目 K,初始化 K 个簇中心的位置,并初始化一个隶属度矩阵,该矩阵定义了每个数据点对各个簇的初始隶属度。

步骤 2:更新隶属度。在每一次迭代中,根据当前各个簇中心的位置,重新计算每个数据点对每个簇中心的隶属度。隶属度的计算通常基于数据点与簇中心之间的距离,距离越近,对该簇的隶属度越高[11]。

步骤 3:更新簇中心。利用当前迭代中所有数据点对各个簇的隶属度作为权重,重新计算每个簇的中心位置。新的簇中心通常是所有数据点的加权平均,权重即为其对该簇的隶属度。

步骤 4:迭代与收敛判断。重复步骤 2 和步骤 3,不断迭代优化[12]。当簇中心位置的变化或预设的目标函数的变化小于某个设定阈值时,算法视为收敛,迭代停止,输出最终的簇中心矩阵和隶属度矩阵。

根据上文内容,将通过标准指数与综合指数划分的数据类别为 X,则 $X=X_1,X_2,\cdots$,以 X_1 为例,则 X_1 的下属数据可以表示为 $\{x_{11},x_{12},\cdots\}$,对于 x_i,对其进行迭代,将数据的迭代过程作为其初始空间的非线性映射过程。映射中,将样本之间的欧氏距离作为参照,对迭代过程描述的方法如式(3)所示[13]。

$$Y=\sum_{i>1}^{X}\sum_{j>1}^{l}\gamma_{ij}^{m}\left\|\varphi(x_j)-\varphi(x_i)\right\|^2 \tag{3}$$

式中：Y 为多类型自然资源现状空间数据迭代处理；l 为最大迭代次数；j 为第 j 个数据；i 为第 i 个数据；γ 为模糊权值系数；m 为核参数；φ 为核中心的中心位置；x_j、x_i 为类别中的对应下属数据。

完成迭代后，利用模糊聚类中的核模糊 C 均值（kernel fuzzy C-means，KFCM）算法。通过设定空间中心函数，将数据映射到高维空间。此过程旨在数据表征的复杂度与模型学习能力之间寻找最佳平衡。利用最大间隔超平面思想对数据进行线性划分，可将构建最优超平面的问题转化为一个带约束的最优化问题[14]。通过设置约束条件，在超平面空间内完成数据的最终模糊聚类。此过程计算方法如式（4）和式（5）所示[15]。

$$\begin{cases}\min \dfrac{1}{2}\|\boldsymbol{W}\|^2 \\ y_i(\boldsymbol{W}x_i+b)\geqslant 1,\ i=1,2,\cdots\end{cases} \tag{4}$$

$$\max(Z)=\sum\alpha_i-\min\frac{1}{2}\|W_i\|^2\beta_i\beta_j x_i x_j \tag{5}$$

式中：$\boldsymbol{W}$ 为超平面法向量；b 为阈值参数；$\max(Z)$ 为处理后的数据；α_i 为拉格朗日乘子；W_i 为维度分量；β 为拉格朗日系数[16]。在式（4）的约束下，对式（5）进行求解，输出对应的数据集合，以此种方式，完成基于迭代模糊聚类的数据处理。

1.3　基于数据相关性的聚类处理

完成上述设计后，利用处理后的数据作为样本，构造全域自然资源信息矩阵，用该矩阵描述数据对象和类之间的关系[17]。设定在数据集中，用不同的减法聚类模糊 C 均值算法（subtractive fuzzy C-means，S-FCM）进行运算，可以获得 H 个存在差异的结果。以此为依据，将 S-FCM 用作基本集成算法，用 S-FCM 提供的半监督信息，将所得到的结果直接表达成类和数据的相关性。在此基础上，将该集成信息矩阵计算方法如式（6）所示。

$$\boldsymbol{A}(H)=[\chi_1,\chi_2,\cdots,\chi_H] \tag{6}$$

式中：$\boldsymbol{A}$ 为集成信息矩阵；χ_H 为信息矩阵差异性结果表达。在基本分簇中，数据集生成多个类数目，则在分簇过程中生成类的总数为多个类数目的总和[18]。因此，将上述构建的信息矩阵进行整合，其矩阵中的第 1 行第 1 列代表第 1 个数据对的第 1 个隶属度。为了获得数据对象间的相似度，需要对同一类中的两个数据进行集成簇的隶属程度定义，此过程计算方法如式（7）所示。

$$L=\boldsymbol{A}(H)(\delta_1,\delta_2) \tag{7}$$

式中：L 为隶属度函数；δ_1、δ_2 为数据之间的隶属度。根据数据之间的隶属关系，进行对应数据的标记，通过对标记信息的聚类，实现对同类属性数据的集成[19]。此过程计算方法如式（8）所示。

$$Q_{ij}=\begin{cases}1, & (x_i,x_j)\in L_1\\ 0, & (x_i,x_j)\in L_2\\ Q_{ij}, & \text{其他}\end{cases} \tag{8}$$

式中：Q_{ij} 为属性数据的集成结果；L_1、L_2 为数据差异性隶属度。在 (x_i,x_j) 对应的数据集合为同类时，划分为同属性数据进行集成，反之，按照另一隶属度划分为不同属性数据进行集成[20]。如果发现数据存在异常或不属于现有类别，筛选出此类数据，启动人工复核修正流程：首先，由专业技术人员对筛选出的异常数据进行逐一审查，结合原始数据来源、实地核查记录及多源数据比对结果，判断其是否为真实异常或误判；其次，针对可修正的异常数据，由人工依据标准化规范进行修正并记录修正原因，对于无法修正或确认为错误的数据，予以剔除并标注剔除依据；最后，所有人工复核修正结果需经二级审核确认，确保处理过程的规范性与结果的可追溯性。通过上述人工复核修正流程，有效提升异常数据处理的准确性与可靠性。通过上述步骤，完成基于数据相关性的聚类处理，实现基于迭代模糊聚类多类型自然资源现状空间数据集成。

2　实验与结果

2.1　实验准备

为验证本文提出的数据集成方法在处理多源异构自然资源空间数据时的有效性，选取试点地区通过“浮标定点监测-船载快速监测-航空遥感-卫星遥感”综合技术手段获取的多类代表性数据作为实验样本。试点地区覆盖范围广泛，涵盖多样化的地理和生态类型，以确保样本数据的代表性和全面性。这些数据主要聚焦于可遥感获取或已有数据库支持的空间信息，包括土地利用/覆被类型、植被覆盖度指数、地表水热状况等遥感衍生数据，以及部分区域的土壤类型和数字高程模型数据。在数据预处理阶段，所有原始数据均经过严格的格式标准化处理，对于栅格数据，根据分析需求进行适当的重采样和格式转换，但并未统一转换为矢量数据，而是保留了其原始格式以充分利用其空间连续性信息；同时，进行坐标统一与质量检查，确保数据的空间参考一致性和质量可靠性，随后存入实验数据库，构成后续集成处理的原始样本。样本数据的详细类别、属性数量及数据条数如表 1 所示，这些数据集在维度和复杂度上具有差异性，具体涵盖了土地利用类型、植被覆盖度、土壤类型与质地、地形地

貌特征、气候数据、水文数据、生物多样性数据及矿产资源分布等多个方面,用于全面测试方法的集成性能。表 1 中详细列出了各数据集的编号、分类数量、属性个数及样本数量,其中试点区域内所有符合分类标准且质量合格的数据均被纳入样本数据集,以确保集成结果的全面性和准确性。

表 1　采样数据类别与规模

数据集编号	分类数量/类	属性/个	样本数量/条
1	4	5	2 015
2	9	12	6 172
3	5	6	5 812
4	4	5	2 551
5	15	23	1 026
6	10	15	384
7	19	25	5 412
8	23	55	6 112

由表 1 可知,针对采集的样本数据,应按数据规范将其转化为标准格式存储在数据中心数据库中,并利用专用资料输入工具进行数据清洗,确保集成前的数据完整性和一致性。

在此基础上,按照本文设计的方法与集成规则进行多类型自然资源现状空间数据集成。为验证本文提出方法的有效性,选择文献[1]和文献[2]中提出的数据集成方法作为对照,参与本次数据集成测试。这两种方法在数据处理领域具有一定的代表性,且均曾被应用于类似的多源异构数据集成场景。

2.2　数据集成归一化处理效果

在自然资源现状数据集成中,归一化处理是关键环节,直接影响后续处理与分析的精度与效果。由于数据样本来源多样化,全域全要素数据的维度存在差异,数据在空间上分布不均匀。为确保实验的公平性,采用与文献[1]处理工具和平台相兼容的通用归一化策略,以确保数据能够在相同的数值范围内进行比较;对于文献[2],则严格按照其描述的最大最小值归一化法进行数据处理,即通过线性变换将原始数据映射到[0,1]的区间内,以消除不同量纲对数据分析结果的影响。因此,应用本文方法、文献[1]方法、文献[2]方法,将各特征点映射到相同的数值区域中,使各类型数据呈现均匀特征。以此为依据,检验数据集成归一化处理效果,其结果如图 2 所示。

通过分析图 2 实验结果并结合自然资源现状分析需求,可知本文设计的基于迭代模糊聚类算法的

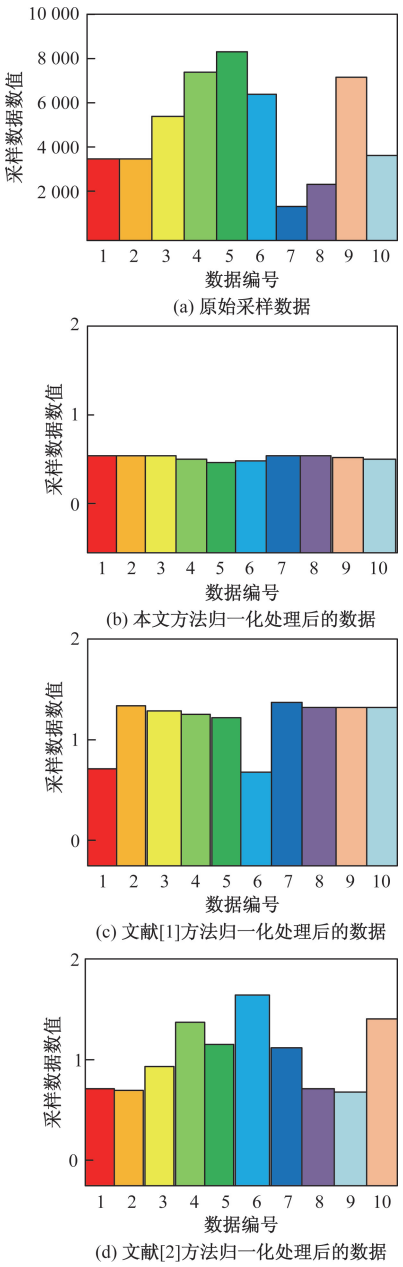


图 2　三种方法数据集成归一化处理效果

方法在全域全要素数据集成中优势显著。该方法不仅使随机分布的采样点呈现规律性,还成功将数据映射至[0,1]的标准化范围,为资源评估、规划和管理提供精准数据基础。其核心在于迭代模糊聚类算法通过合理设定抽样范围与模糊权值系数,减少了训练样本量,提高了处理器效率与准确率,同时基于数据相关性的聚类处理流程增强了数据在高维空间的映射能力。文献[1]和文献[2]方法虽能实现数据分布的均匀化处理,但在数据映射的完整性和分析精度上仍有提升空间,进一步凸显了本文方法在处理复杂自然资源数据时的优越性。

2.3　集成数据类别相似性检验

2.3.1　同一样本测试数据集成相似性检验

完成上述测试后，为了检验集成后的多类型自然资源现状空间数据集是否符合类别，使用相似度测量方法，进行集成后数据的集中测试。如果发现集成后的数据与聚类中心具有较高相似度，说明在集成过程中没有产生显著的数据类混淆现象。反之，说明样本数据存在混淆与异常现象。本次实验以同一样本集合的自然资源数据为例，对其进行相似度分析，结果如图 3 所示。

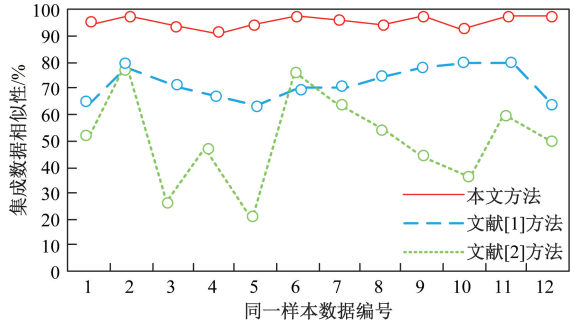


图 3　同一样本测试数据集成相似性检验结果

采用本文设计的基于迭代模糊聚类算法的方法进行同一样本测试数据集成后，样本编码数据与聚类中心的相似度超过了 90%。这一显著的高相似度，主要得益于迭代模糊聚类算法能够精准地捕捉数据间的内在关联性和类别特征。该算法通过不断优化聚类中心，使得集成后的数据更加贴近其真实的类别归属，从而在自然资源现状分析中显著提升了数据集成的准确性和有效性。这对于准确识别自然资源分布、评估资源利用状态以及制定科学合理的资源管理策略至关重要。在此基础上，分析文献[1]和文献[2]方法的应用效果可以看出，对应方法集成后的样本数据与聚类中心的相似度在 60%～80%与 20%～80%，集成效果均略差于本文方法。

2.3.2　多样本测试数据集成相似性检验

为避免实验结果的偶然性，选择多个数据样本，参照上述方式，对多样本测试数据集成相似性进行检验，其结果如图 4 所示。

在相同的测试条件下，本方法多次测试均保持了集成样本与聚类中心的高度相似性。这一优异表现，主要得益于迭代模糊聚类算法与数据相关性聚类处理流程的紧密结合。该算法能够动态调整聚类中心和模糊权值系数，精准捕捉自然资源数据间的复杂关联性和类别特征，从而在自然资源现状

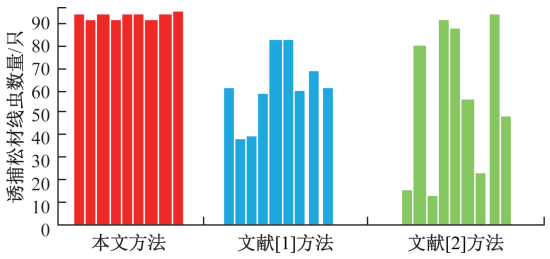


图 4　多样本测试数据集成相似性检验结果

分析中保持较高的聚类准确性和稳定性。这对于准确分析自然资源分布、评估资源利用效率及制定科学管理策略具有重要意义。文献[1]和文献[2]方法虽然在一定程度上实现了数据集成，但在处理复杂自然资源数据时，其捕捉数据间复杂关联性和保持聚类稳定性的能力略显不足，从而凸显了本文方法在提升数据集成质量和分析精度上的显著优势。

2.4　与经典地质解译方法的混淆矩阵对比

为进一步验证本方法在自然资源现状数据集成中的类别区分能力，选取经典地质解译方法中的 CUI 指数法作为对比基准。采用混淆矩阵对本方法与综合城市化指数(comprehensive urbanization index，CUI)法的分类结果进行定量评估，对比两者在多类型自然资源数据分类中的总体精度、Kappa 系数及各类别的用户精度、生产者精度。

实验选取表 1 中具有明确地质解译背景的数据集，以人工解译验证结果作为真实参考。本文方法按上文流程执行分类集成；CUI 指数法按照其经典流程，计算各样本单元的 CUI 指数后依据经验阈值进行分类。混淆矩阵对比结果如表 2 所示。

本文方法在总体精度和 Kappa 系数上均显著优于 CUI 指数法，表明本文方法具有更高的分类一致性与可靠性。在类别平均用户精度与生产者精度方面，本文方法同样保持明显优势，分别高出 CUI 指数法 12.3 个百分点和 11.8 个百分点。这一结果主要得益于本文提出的迭代模糊聚类算法能够有效处理自然资源数据中普遍存在的边界模糊与属性重叠问题，而 CUI 指数法基于固定阈值的硬性划分方式难以准确捕捉此类复杂边界特征。

表 2　本文方法与 CUI 指数法的混淆矩阵对比结果

评估指标	本文方法	CUI 指数
总体精度/%	94.6	82.3
Kappa 系数	0.92	0.75
类别平均用户精度/%	92.8	80.5
类别平均生产者精度/%	93.5	81.7

3 讨论

本文提出的基于迭代模糊聚类的数据集成方法,通过构建“标准指数-综合指数”分类体系并结合数据相关性聚类处理,显著提升了数据集成的内在一致性与类别纯净度。实验结果表明,该方法在数据归一化与类别相似性等关键指标上均优于传统方法,有效解决了现有研究在应对复杂自然资源数据时面临的集成效果不佳与效率低下等核心问题。

然而,在深入讨论自然资源现状数据处理的具体应用时,论文中对于数据类型及其对应处理方法的阐述不够详尽,且本文方法对初始参数设置较为敏感。

针对参数敏感性问题,未来可探索自适应参数优化策略,如引入基于数据特性的动态模糊权值调整机制,根据数据分布特征自动优化聚类过程中的关键参数,减少人工调参的盲目性,从而提升方法的鲁棒性与自动化水平。

同时,通过细化自然资源现状数据的分类及其适用性分析,土地利用类型数据因其空间分布明确、类别边界清晰,适合采用硬聚类方法进行初步分类,再结合本文提出的迭代模糊聚类算法进行精细化处理,以提高分类精度。植被覆盖度数据则因其存在过渡区域和模糊边界,直接应用迭代模糊聚类算法能够更好地捕捉其内在的模糊性,实现更准确的分类。土壤类型与质地数据、地形地貌特征数据等,由于其属性复杂且存在重叠,同样适合采用本文方法进行处理,以揭示其深层次的类别特征和关联性。

在应用场景方面,本文提出的方法具有广泛的适用性。在遥感地质图像分析领域,该方法能够智能分类遥感图像,有效划分地质图中的过渡性区域,提升解译准确性,为地质勘探和资源开发提供有力支持。在地质数据集成与成矿预测方面,通过对多源地质、地球物理、地球化学数据的聚类分析,可以识别相似成矿背景的地质单元,为圈定成矿远景区提供科学依据。此外,在资源调查与监测、生态保护与修复、国土空间规划等多个领域,本文方法均能通过生成高质量的自然资源信息矩阵,为资源评估、规划和管理提供精准数据基础,显著减少数据混淆和异常值的影响,提高资源调查的准确性和效率。

针对大规模数据处理效率提升问题,未来可进一步探讨分布式计算框架的具体实现路径,例如,采用 MapReduce 或 Spark 等分布式计算模型,将数据集成任务分解为多个子任务并行处理,充分利用集群计算资源,显著提高数据处理速度和可扩展性。同时,结合深度学习技术优化聚类算法的性能,通过构建高效的神经网络模型自动提取数据特征,减少人工特征工程的复杂性,从而进一步提升大规模自然资源数据集成的效率和精度。

综上所述,通过细化自然资源现状数据的分类及其适用性分析,明确方法在多个应用场景中的具体价值,并补充自适应参数优化策略及分布式计算等具体实现路径的探讨,不仅增强了讨论章节的支撑力,也为实际应用提供了更具操作性的参考。

未来,将继续探索自适应参数优化策略以提升方法的鲁棒性与自动化水平,同时研究分布式计算框架与深度学习技术的融合,以应对自然资源大数据集成日益增长的挑战。

4 结论

针对自然资源现状数据集成中存在的类别混淆与规范性不足等核心问题,本文提出一种基于迭代模糊聚类的数据集成方法。构建“标准指数-综合指数”分类体系,为多源异构的自然资源数据提供了统一的分类依据;引入迭代模糊聚类算法对已分类数据进行高效处理,通过优化模糊权值与迭代过程,在减少训练样本量的同时提升了聚类精度与运算效率;设计基于数据相关性的聚类处理流程,通过构造全域自然资源信息矩阵与定义数据隶属关系,实现了同类数据的高质量集成与异常数据的有效筛选。实验验证表明,该方法在数据集成归一化处理与类别相似性检验方面均优于传统方法,能够生成规律性强、同类别相似度高的数据,有效解决现有研究在应对复杂自然资源数据时面临的集成效果不佳与效率低下的问题,为构建精准高效的自然资源业务信息化体系提供了可靠的方法支撑。然而,本文方法对初始参数设置较为敏感,且在超大规模数据集下的计算效率仍有优化空间。未来,将探索自适应参数优化策略以提升方法的鲁棒性与自动化水平,并研究分布式计算框架与深度学习技术的融合,以期更好地应对自然资源大数据集成日益增长的挑战。

参考文献

[1] 王世杰,李一鸣,植殷,等. 小样本数据下基于 K-Means 聚类和集成学习的混凝投药预测[J]. 环境工程学报,2024,18(1):181-188.

[2] 文敏,月一,张怀东,等. 地质调查业务管理数据集成辅助决策系统架构与关键技术[J]. 地质通报,2024,43

(7)：1221-1232.
[3] 孙中彬，刁宇轩，马苏洋. 基于安全欠采样的不均衡多标签数据集成学习方法[J]. 电子学报，2024，52(10)：3392-3408.
[4] 王江江，邓洪达，刘艺，等. 数据中心综合能源系统配置与运行的集成优化[J]. 科学技术与工程，2023，23(5)：1968-1977.
[5] 何帆，赵秀翔，何梦旗，等. 多品种小批量离散制造过程多源异构数据集成与智能运维方法研究[J]. 现代制造工程，2024(10)：49-59.
[6] 王辉. 一种基于特征类的自然资源时空数据治理方法[J]. 住宅与房地产，2026(6)：25-27.
[7] 王启飞，王俊龙，刘昊霖，等. 煤矿安全风险分析的文本数据模型与集成分析平台[J]. 安全与环境学报，2024，24(11)：4358-4365.
[8] 邹建军. 地理信息大数据在自然资源资产核算中的应用[J]. 中国资源综合利用，2026，44(2)：89-91.
[9] 潘永淇，刘毅，程子文，等. 面向疫情防控政策生成的区块链数据集成系统[J]. 计算机工程与设计，2023，44(11)：3469-3476.
[10] 刘信冰. 土地信息技术在自然资源管理中的应用[J]. 农村科学实验，2026(4)：40-42.
[11] 仇巍巍，吴初国，齐书花，等. 实景三维与元宇宙及其在自然资源领域的应用探究[J]. 自然资源信息化，2024(2)：1-6.
[12] 邹瑶，陈韵，徐杰. 规划与自然资源主数据模型的建立方法[J]. 北京测绘，2025，39(4)：490-495.
[13] 季玉洁，李观石，唐天琪. 基于Echarts的自然资源大数据可视化系统设计与实现[J]. 无线互联科技，2025，22(7)：44-51.
[14] 蔡文博，王薇，朱青，等. 基于谷歌地球引擎大数据支撑的自然资源生态安全应用研究进展[J]. 生态学报，2025，45(7)：3544-3554.
[15] 张可心，郭振华，顾巍巍. 基于FME的自然资源确权数据自动化处理[J]. 经纬天地，2025(1)：95-100.
[16] 刘爽. 基于关联规则算法的自然资源数据集成存储方法[J]. 电脑编程技巧与维护，2025(11)：100-102.
[17] 冯亚林. 基于多源测绘数据的自然资源综合调查监测体系构建[J]. 生态与资源，2025(12)：31-33.
[18] 赵少杰，牟晓慧，刘辉. 基于多源地理信息数据的自然资源调查系统构建研究[J]. 科技资讯，2025，23(21)：22-24.
[19] 周洪保. 基于实景三维的自然资源调查成果可视化监管研究[J]. 资源导刊，2025(24)：27-30.
[20] 唐世川，邓卫. 自然资源调查监测数据自动化处理研究[J]. 测绘与空间地理信息，2025，48(10)：80-84.

Data Integration of Natural Resource Status Based on Iterative Fuzzy Clustering

WANG Xingjuan，JI Xiang
(Jining Center of Land Spatial Data and Remote Sensing Technology，Jining 272000，China)

Abstract：In order to effectively solve the key problems faced in the integration process of natural resource status data，such as category confusion，insufficient standardization and low processing efficiency，a natural resource status data integration method based on iterative fuzzy clustering was proposed. By using the standard index method and the "Nemero" comprehensive index method，a "standard index comprehensive index" grading system was constructed to scientifically classify the current status data of natural resources in the entire region and all factors. Furthermore，the iterative fuzzy clustering algorithm was introduced to centrally process the classified data. By optimizing the fuzzy weights and iterative process，the training sample size was significantly reduced while improving the training efficiency and clustering accuracy of the processor. On this basis，a global natural resource information matrix was constructed using processed data，and a clustering processing flow based on data correlation was designed to achieve high-precision integration of similar attribute data and effective screening of abnormal data. The experimental results show that this method performs well in data integration normalization processing，with significant regularity and high similarity between categories in the processed data. It effectively meets the practical needs of natural resource status analysis and provides methodological support for building a precise and efficient all-factor information system for natural resource business.

Keywords：iterative fuzzy clustering；integrated methods；current situation data；natural resources；double level classification system