



(12) 发明专利申请

(10) 申请公布号 CN 117271683 A

(43) 申请公布日 2023. 12. 22

(21) 申请号 202311174337.X

G06F 18/2433 (2023.01)

(22) 申请日 2023.09.13

G06F 18/23 (2023.01)

(71) 申请人 北京欣智恒科技股份有限公司
地址 100080 北京市海淀区海淀南路19号
时代网络大厦8000室

(72) 发明人 张杏芝 牛子宁 冀光 张卫
付佰兴 刘喜平 荆美佳

(74) 专利代理机构 北京同春专利代理事务所
(普通合伙) 16239
专利代理师 朱怡蔓

(51) Int. Cl.

G06F 16/29 (2019.01)

G01C 15/00 (2006.01)

G06F 18/10 (2023.01)

G06F 18/24 (2023.01)

权利要求书3页 说明书15页 附图3页

(54) 发明名称

一种测绘数据的智能分析评价方法

(57) 摘要

本发明公开了一种测绘数据智能分析评价方法,涉及测绘数据信息处理技术领域,解决了传统测绘数据分析评价方法中数据分析不全面、数据质量评价不精确和异常检测方面存在的缺点;采用的方法包括:(S1)通过数据预处理方法对测绘数据进行清洗去噪、校正转换和缺省值填充操作;(S2)通过图像处理工具从测绘数据中提取特征,并通过特征选择器排除冗余和不相关的特征;(S3)通过空间分析模块对地理信息进行空间叠加;(S4)通过统计分析模块对测绘数据进行统计描述、统计推断和假设检验处理;(S5)通过智能分析模块对测绘数据进行分类、预测和优化处理;(S6)通过智能评价模块进行数据评价与异常检测;(S7)通过可视化模块将评价内容和分析后的测绘数据进行展示。



1. 一种测绘数据智能分析评价方法,其特征在于:包括以下步骤:

步骤1、通过数据预处理方法对测绘数据进行清洗去噪、校正转换和缺省值填充操作;所述校正转换通过坐标校正和投影转换方法以确保数据的空间准确性和一致性;所述缺省值填充通过插值和统计方法填充缺失的数据值;

步骤2、通过图像处理工具从测绘数据中提取特征,并通过特征选择器排除冗余和不相关的特征,以提高后续分析的效率和准确性;

步骤3、通过空间分析模块对地理信息进行空间叠加,所述空间分析模块包括邻域分析单元、聚类分析单元和空间关联性分析单元;所述邻域分析单元通过空间索引结构和欧式距离评估地物的邻域密度和距离指标,以确定地物的集聚程度;所述聚类分析单元通过聚类算法识别具有相似空间特征的地物群集;所述空间关联性分析单元通过空间自相关统计方法衡量地物之间的空间相关性;

步骤4、通过统计分析模块对测绘数据进行统计描述、统计推断和假设检验处理,以揭示测绘数据的内在规律和特征;所述统计分析模块通过计算均值、方差和百分位数以描述测绘数据特征,并通过相关系数度量不同变量之间的相关程度;

步骤5、通过智能分析模块对测绘数据进行分类、预测和优化处理,以提高测绘数据的质量和效率;

步骤6、通过智能评价模块进行数据评价与异常检测;所述智能评价模块包括数据质量评价单元和异常检测单元;所述数据质量评价单元通过数据挖掘方法和质量评价模型对测绘数据的准确性、完整性和一致性进行评估;所述异常检测单元通过数据纠错引擎检查并修复测绘数据中的错误、异常值和离群点,以提高数据的可靠性和准确性;

步骤7、通过可视化模块将评价内容和分析后的测绘数据进行展示。

2. 根据权利要求1所述的一种测绘数据智能分析评价方法,其特征在于:所述图像处理工具包括特征提取模块、图像分割模块、特征降维模块和后处理模块;所述特征提取模块通过局部二值模式和梯度方向直方图计算和提取特征;所述图像分割模块包括阈值分割单元和边缘分割单元;所述阈值分割单元根据图像亮度或颜色的阈值,通过基于颜色空间的阈值分割方法将图像分为不同的区域;所述边缘分割单元通过索贝尔算子检测图像中物体边缘进行分割,以提取出物体的边界信息;所述特征降维模块包括特征选择单元和特征降维单元;所述特征选择单元通过统计方法和信息论方法评估特征的重要性和相关性,以选取特征;所述特征降维单元通过主成分提取和线性判别分析将高维的特征空间降低到低维空间,以降低特征复杂度;所述后处理模块包括噪声去除单元和目标合并单元;所述噪声去除单元通过滤波器对分类和识别结果中的噪声进行去除,以提高结果的准确性;所述目标合并单元通过连通区域分析方法将相邻的目标进行合并,以形成完整目标区域;所述特征提取模块的输出端与所述图像分割模块的输入端连接;所述图像分割模块的输出端与所述特征降维模块的输入端连接;所述特征降维模块的输出端与所述后处理模块的输入端连接。

3. 根据权利要求1所述的一种测绘数据智能分析评价方法,其特征在于:所述邻域分析单元通过邻域密度指标对地物的邻域密度进行分析,并通过R树对高维数据和多边形对象进行数据化分析;其中,领域分析的公式表达式为:

$$K = \log \sum e^y \sqrt{y + q} - F \frac{-q \pm \sqrt{F^2 - 4r}}{2e^y} + r \quad (1)$$

在公式(1)中,K表示邻域密度,y表示邻域内点的数量,r表示邻域半径,q表示邻域内线段的长度,F表示邻域内面积;

所述聚类分析单元通过迭代实现最小化总体误差;迭代步骤为:

- (S1) 初始化聚类中心;
- (S2) 计算样本与聚类中心的距离,将样本划分到最近的类别;
- (S3) 更新聚类中心为每个类别的均值;
- (S4) 重复步骤(S2)和(S3),直到满足停止条件;

所述空间关联性分析单元通过莫兰指数衡量空间数据分布的集聚程度,莫兰指数的公式表达式为:

$$I = \frac{W}{N} \prod \sec^{-1} W + \log \sqrt{i^2 + j^2} - \cos N (j - i)^2 \quad (2)$$

在公式(2)中,N表示样本点数量,W表示空间邻接矩阵,i表示空间距离权重参数,j表示参数对应的属性值;另外,空间关联性分析单元通过吉尔利指数衡量空间数据的离散程度,吉尔利指数的公式表达式为:

$$C = \frac{W}{N} \sum \sec^{-1} (j - i) + \frac{\sqrt{i^2 - 4j}}{M} - W \log \frac{\pi}{2} \sqrt{i^2 + j^2} - N \cos N^2 \quad (3)$$

在公式(3)中,N表示样本点数量,W表示空间邻接矩阵,i表示空间距离权重参数,j表示参数对应的属性值,M表示所有样本点的属性值均值;吉尔利指数的取值范围为[0,2],当值接近0时,表示正相关性,即相似值与相似值相邻;当值接近2时,表示负相关性,即相似值与不相似值相邻;当值接近1时,表示无空间自相关性。

4. 根据权利要求1所述的一种测绘数据智能分析评价方法,其特征在于:所述统计分析模块包括统计描述单元、统计推断单元和假设检验处理单元;所述统计描述单元通过描述性统计和极差分析描述和总结测绘数据的特征和分布情况;所述描述性统计通过数据的均值、中位数、众数和标准差获得数据中心趋势和离散程度,以描述数据的集中程度和分散程度;所述极差分析通过数据最大值和最小值之间的差异衡量数据的变异程度;所述统计推断单元基于从测绘数据中获取的样本数据,通过参数估计和置信区间估计方法推断总体的性质和特征;所述假设检验处理单元通过集计分析、相关分析和回归分析验证或否定在测绘数据分析中提出的假设;所述集计分析通过交叉表分析和频数分布分析方法了解数据要素之间的关系;所述相关分析通过皮尔逊相关系数判断数据之间的线性关系强度和方向;所述回归分析通过线性回归函数关系模型预测未知数据。

5. 根据权利要求1所述的一种测绘数据智能分析评价方法,其特征在于:所述智能分析模块包括特征工程单元和模型训练评估单元;所述特征工程单元通过特征选择和特征转换的方式对测绘数据进行预处理,以提取特征;所述特征选择通过信息增益衡量特征对分类问题的贡献度;并通过相关系数计算特征与目标变量间的相关性,以提取目标特征;所述特征转换通过主成分提取和线性判别分析将原始特征转换为新的特征表示,以提高模型性能;所述主成分提取通过将多维数据降维为低维度表示,以消除冗余信息,提取关键特征;所述线性判别分析通过将高维数据投影到低维空间,并保留区分不同类别的信息,以实现分类问题的优化和提高;所述模型训练评估单元通过随机森林算法训练模型,并通过评估指标对模型的性能进行评估,所述评估指标包括准确率、召回率和调和平均数。

6. 根据权利要求1所述的一种测绘数据智能分析评价方法,其特征在于:所述可视化模

块包括可视化展示单元和数据整合单元;所述数据整合单元通过数据集成分析方法将空间分析模块、统计分析模块和智能分析模块的分析结果进行整合,以得出综合评价结果,最终通过可视化展示单元进行展示;所述可视化展示单元通过地理信息系统和数据可视化工具,将分析结果以地图或图表形式展示,以帮助用户理解和使用测绘数据;并通过报告生成器根据分析结果生成文档报告。

7. 根据权利要求1所述的一种测绘数据智能分析评价方法,其特征在于:所述数据挖掘方法的工作方式步骤为:

(S1) 通过测绘数据采集设备获取测绘数据;

(S2) 通过数据清洗工具进行去除重复值、处理缺失值和处理异常值操作,以确保数据的准确性和完整性;

(S3) 通过特征选择算法从原始数据中选择与评估目标相关的特征;所述特征选择算法包括信息增益、相关系数提取和主成分提取操作;

(S4) 通过特征转换方法将选择的特征进行转换,以便于后续分析和建模;

(S5) 通过关联规则挖掘算法,构建评估模型;

(S6) 使用已标注的数据集对构建的模型进行评估,比较模型预测结果与实际情况的差异,并计算评估指标;

(S7) 根据评估结果,对测绘数据的准确性、完整性和一致性进行解释和分析,发现数据的潜在问题和规律,并提出改进建议。

8. 根据权利要求1所述的一种测绘数据智能分析评价方法,其特征在于:所述质量评价模型包括数据准确性评估模块、数据完整性评估模块和数据一致性评估模块;所述数据准确性评估模块包括空间一致性评估单元和属性一致性评估单元;所述空间一致性评估单元通过重叠度计算和误差椭圆生成方法比较同一区域内不同数据源采集的相同地理要素,检测数据之间的几何一致性;所述属性一致性评估单元通过属性值比较和相似度计算方法比较同一地理要素在不同数据源中的属性值,以评估数据的属性准确性;所述数据完整性评估模块包括缺失值检测单元和异常值检测单元;所述缺失值检测单元通过缺失值比例统计和缺失机制分析的方式分析数据中的缺失情况,识别缺失数据的位置和类型;以评估数据完整性;所述异常值检测单元通过离群点检测和异常值分析方法检测数据中的异常值,以识别数据错误或异常情况;所述数据一致性评估模块包括逻辑一致性检测单元和时间一致性检测单元;所述逻辑一致性检测单元通过拓扑关系分析和拓扑错误检查方法分析数据之间的逻辑关系,检测数据是否符合预期的一致性规则,以评估数据的一致性;对于具有时间属性的数据,通过时间一致性检测单元评估数据一致性;所述时间一致性检测单元通过时间序列分析和时间窗口比较方法检测数据在时间上的一致性。

9. 根据权利要求1所述的一种测绘数据智能分析评价方法,其特征在于:所述数据纠错引擎包括数据预处理单元、异常识别单元、异常修复单元和数据验证单元;所述数据预处理单元通过删除重复值、处理缺失值以及处理格式错误的方式保证数据的完整性和一致性;所述异常识别单元通过箱线图和孤立森林算法识别测绘数据中的异常值和离群点;所述异常修复单元通过线性插值、多项式插值、回归模型拟合和基于规则的替换方法修复测绘数据中的错误、异常值和离群点;所述数据验证单元通过数据一致性检查和数据分布分析方法确保修复后的数据的准确性和可靠性。

一种测绘数据的智能分析评价方法

技术领域

[0001] 本发明涉及测绘数据信息处理技术领域,且更具体地涉及一种测绘数据智能分析评价方法。

背景技术

[0002] 测绘数据是指通过测绘技术获取的地理信息数据,广泛应用于土地规划、城市建设、资源管理等领域。随着科技的不断进步和信息化的发展,测绘数据的获取和处理能力逐渐提升,为各行各业的发展提供了重要支持。通过对测绘数据进行分析评价,可以识别和纠正数据中的错误、偏差和不一致性,以提高数据的准确性和可信度。这有助于保证数据在后续应用和决策中的有效性。另外,测绘数据的分析评价可以帮助用户更好地理解数据,发现其中的规律和关联,并提供关键的信息和指标用于决策制定。例如,在城市规划中,通过对地理和空间数据进行评价,可以为土地利用、交通规划等方面的决策提供科学依据。测绘数据的分析评价对于确保数据质量、支持决策制定、提高数据应用效果、优化数据采集和处理以及促进数据共享和交互等方面具有重要的意义。在过去的几十年中,测绘技术和数据分析方法经历了多次革新和改进。在早期,测绘数据的处理主要依赖于人工操作和传统的测量设备,数据量较小,分析方法也相对简单。然而,随着测绘数据的规模和复杂性不断增加,传统的分析评价方法逐渐显现出一些缺点和挑战:

[0003] 首先,传统方法往往对地理信息的空间叠加和关联性分析能力有限。传统方法通常只能通过简单的邻域分析来评估地物的集聚程度,无法提供更全面和精细的空间分析结果。其次,传统方法的统计描述和推断能力有一定局限。传统方法通常仅使用基本的统计指标如均值、方差和百分位数来描述测绘数据的基本特征,无法提供更多维度的统计信息。此外,传统方法在推断分析和假设检验处理方面也存在不足,无法揭示测绘数据中的潜在规律和特征,限制了对数据背后因果关系的探究能力。第三,传统方法对测绘数据的分类、预测和优化处理能力有限。传统方法主要依赖人工经验和规则来进行数据分析和处理,缺乏自动化和智能化的机器学习算法支持。因此,传统方法往往无法充分挖掘大规模测绘数据中的潜在模式和关联性,限制了数据分析的准确性和效率。最后,传统方法存在着一定的局限性,无法有效识别和纠正测绘数据中的错误、异常值和离群点,导致数据可靠性和准确性受到威胁。

[0004] 因此,为了解决传统测绘数据分析评价方法中分析效果差和数据质量评价以及异常检测方面存在的缺点和限制,本发明公开一种测绘数据智能分析评价方法。

发明内容

[0005] 针对现有技术的不足,本发明公开了一种测绘数据智能分析评价方法,本发明通过空间分析模块提供了更全面和精细的空间分析能力,揭示了地物的集聚程度和空间相关性;通过统计分析模块全面地了解测绘数据的基本特征和不同变量之间的相关程度;通过智能分析模块对测绘数据进行自动化的分类、预测和优化处理,提高了测绘数据的质量和

效率;通过智能评价模块对测绘数据的准确性、完整性和一致性进行评估,并检查和修复测绘数据中的错误和异常值,提高了数据的可靠性和准确性。

[0006] 为了实现上述技术效果,本发明采用以下技术方案:

[0007] 一种测绘数据智能分析评价方法,其中所述方法包括:

[0008] 作为本发明进一步的技术方案,包括以下步骤:

[0009] 步骤1、通过数据预处理方法对测绘数据进行清洗去噪、校正转换和缺省值填充操作;所述校正转换通过坐标校正和投影转换方法以确保数据的空间准确性和一致性;所述缺省值填充通过插值和统计方法填充缺失的数据值;

[0010] 步骤2、通过图像处理工具从测绘数据中提取特征,并通过特征选择器排除冗余和不相关的特征,以提高后续分析的效率 and 准确性;

[0011] 步骤3、通过空间分析模块对地理信息进行空间叠加,所述空间分析模块包括邻域分析单元、聚类分析单元和空间关联性分析单元;所述邻域分析单元通过空间索引结构和欧式距离评估地物的邻域密度和距离指标,以确定地物的集聚程度;所述聚类分析单元通过聚类算法识别具有相似空间特征的地物群集;所述空间关联性分析单元通过空间自相关统计方法衡量地物之间的空间相关性;

[0012] 步骤4、通过统计分析模块对测绘数据进行统计描述、统计推断和假设检验处理,以揭示测绘数据的内在规律和特征;所述统计分析模块通过计算均值、方差和百分位数以描述测绘数据特征,并通过相关系数度量不同变量之间的相关程度;

[0013] 步骤5、通过智能分析模块对测绘数据进行分类、预测和优化处理,以提高测绘数据的质量和效率,

[0014] 步骤6、通过智能评价模块进行数据评价与异常检测;所述智能评价模块包括数据质量评价单元和异常检测单元;所述数据质量评价单元通过数据挖掘方法和质量评价模型对测绘数据的准确性、完整性和一致性进行评估;所述异常检测单元通过数据纠错引擎检查并修复测绘数据中的错误、异常值和离群点,以提高数据的可靠性和准确性;

[0015] 步骤7、通过可视化模块将评价内容和分析后的测绘数据进行展示。

[0016] 作为本发明进一步的技术方案,所述图像处理工具包括特征提取模块、图像分割模块、特征降维模块和后处理模块;所述特征提取模块通过局部二值模式和梯度方向直方图计算和提取特征;所述图像分割模块包括阈值分割单元和边缘分割单元;所述阈值分割单元根据图像亮度或颜色的阈值,通过基于颜色空间的阈值分割方法将图像分为不同的区域;所述边缘分割单元通过索贝尔算子检测图像中物体边缘进行分割,以提取出物体的边界信息;所述特征降维模块包括特征选择单元和特征降维单元;所述特征选择单元通过统计方法和信息论方法评估特征的重要性和相关性,以选取特征;所述特征降维单元通过主成分提取和线性判别分析将高维的特征空间降低到低维空间,以降低特征复杂度;所述后处理模块包括噪声去除单元和目标合并单元;所述噪声去除单元通过滤波器对分类和识别结果中的噪声进行去除,以提高结果的准确性;所述目标合并单元通过连通区域分析方法将相邻的目标进行合并,以形成完整目标区域;所述特征提取模块的输出端与所述图像分割模块的输入端连接;所述图像分割模块的输出端与所述特征降维模块的输入端连接;所述特征降维模块的输出端与所述后处理模块的输入端连接。

[0017] 作为本发明进一步的技术方案,所述邻域分析单元通过邻域密度指标对地物的邻

域密度进行分析,并通过R树对高维数据和多边形对象进行数据化分析;其中,领域分析的公式表达式为:

$$[0018] \quad K = \log \sum e^y \sqrt{y+q} - F \frac{-q \pm \sqrt{F^2 - 4r}}{2e^y} + r \quad (1)$$

[0019] 在公式(1)中,K表示邻域密度,y表示邻域内点的数量,r表示邻域半径,q表示邻域内线段的长度,F表示邻域内面积;

[0020] 所述聚类分析单元通过迭代实现最小化总体误差;迭代步骤为:

[0021] (S1) 初始化聚类中心;

[0022] (S2) 计算样本与聚类中心的距离,将样本划分到最近的类别;

[0023] (S3) 更新聚类中心为每个类别的均值;

[0024] (S4) 重复步骤(S2)和(S3),直到满足停止条件;

[0025] 所述空间关联性分析单元通过莫兰指数衡量空间数据分布的集聚程度,莫兰指数的公式表达式为:

$$[0026] \quad I = \frac{W}{N} \prod \sec^{-1} W + \log \sqrt{i^2 + j^2} - \cos N (j - i)^2 \quad (2)$$

[0027] 在公式(2)中,N表示样本点数量,W表示空间邻接矩阵,i表示空间距离权重参数,j表示参数对应的属性值;另外,空间关联性分析单元通过吉尔利指数衡量空间数据的离散程度,吉尔利指数的公式表达式为:

$$[0028] \quad C = \frac{W}{N} \sum \sec^{-1} (j - i) + \frac{\sqrt{i^2 - 4j}}{M} - W \log \frac{\pi}{2} \sqrt{i^2 + j^2} - N \cos N^2 \quad (3)$$

[0029] 在公式(3)中,N表示样本点数量,W表示空间邻接矩阵,i表示空间距离权重参数,j表示参数对应的属性值,M表示所有样本点的属性值均值;吉尔利指数的取值范围为[0,2],当值接近0时,表示正相关性,即相似值与相似值相邻;当值接近2时,表示负相关性,即相似值与不相似值相邻;当值接近1时,表示无空间自相关性。

[0030] 作为本发明进一步的技术方案,所述统计分析模块包括统计描述单元、统计推断单元和假设检验处理单元;所述统计描述单元通过描述性统计和极差分析描述和总结测绘数据的特征和分布情况;所述描述性统计通过数据的均值、中位数、众数和标准差获得数据中心趋势和离散程度,以描述数据的集中程度和分散程度;所述极差分析通过数据最大值和最小值之间的差异衡量数据的变异程度;所述统计推断单元基于从测绘数据中获取的样本数据,通过参数估计和置信区间估计方法推断总体的性质和特征;所述假设检验处理单元通过集计分析、相关分析和回归分析验证或否定在测绘数据分析中提出的假设;所述集计分析通过交叉表分析和频数分布分析方法了解数据要素之间的关系;所述相关分析通过皮尔逊相关系数判断数据之间的线性关系强度和方向;所述回归分析通过线性回归函数关系模型预测未知数据。

[0031] 作为本发明进一步的技术方案,所述智能分析模块包括特征工程单元和模型训练评估单元;所述特征工程单元通过特征选择和特征转换的方式对测绘数据进行预处理,以提取特征;所述特征选择通过信息增益衡量特征对分类问题的贡献度;并通过相关系数计算特征与目标变量间的相关性,以提取目标特征;所述特征转换通过主成分提取和线性判别分析将原始特征转换为新的特征表示,以提高模型性能;所述主成分提取通过将多维数据降维为低维度表示,以消除冗余信息,提取关键特征;所述线性判别分析通过将高维数据

投影到低维空间,并保留区分不同类别的信息,以实现分类问题的优化和提高;所述模型训练评估单元通过随机森林算法训练模型,并通过评估指标对模型的性能进行评估,所述评估指标包括准确率、召回率和调和平均数。

[0032] 作为本发明进一步的技术方案,所述可视化模块包括可视化展示单元和数据整合单元;所述数据整合单元通过数据集成分析方法将空间分析模块、统计分析模块和智能分析模块的分析结果进行整合,以得出综合评价结果,最终通过可视化展示单元进行展示;所述可视化展示单元通过地理信息系统和数据可视化工具,将分析结果以地图或图表形式展示,以帮助用户理解和使用测绘数据;并通过报告生成器根据分析结果生成文档报告。

[0033] 作为本发明进一步的技术方案,所述数据挖掘方法的工作方式步骤为:

[0034] (S1) 通过测绘数据采集设备获取测绘数据;

[0035] (S2) 通过数据清洗工具进行去除重复值、处理缺失值和处理异常值操作,以确保数据的准确性和完整性;

[0036] (S3) 通过特征选择算法从原始数据中选择与评估目标相关的特征;所述特征选择算法包括信息增益、相关系数提取和主成分提取操作;

[0037] (S4) 通过特征转换方法将选择的特征进行转换,以便于后续分析和建模;

[0038] (S5) 通过关联规则挖掘算法,构建评估模型;

[0039] (S6) 使用已标注的数据集对构建的模型进行评估,比较模型预测结果与实际情况的差异,并计算评估指标;

[0040] (S7) 根据评估结果,对测绘数据的准确性、完整性和一致性进行解释和分析,发现数据的潜在问题和规律,并提出改进建议。

[0041] 作为本发明进一步的技术方案,所述质量评价模型包括数据准确性评估模块、数据完整性评估模块和数据一致性评估模块;所述数据准确性评估模块包括空间一致性评估单元和属性一致性评估单元;所述空间一致性评估单元通过重叠度计算和误差椭圆生成方法比较同一区域内不同数据源采集的相同地理要素,检测数据之间的几何一致性;所述属性一致性评估单元通过属性值比较和相似度计算方法比较同一地理要素在不同数据源中的属性值,以评估数据的属性准确性;所述数据完整性评估模块包括缺失值检测单元和异常值检测单元;所述缺失值检测单元通过缺失值比例统计和缺失机制分析的方式分析数据中的缺失情况,识别缺失数据的位置和类型;以评估数据完整性;所述异常值检测单元通过离群点检测和异常值分析方法检测数据中的异常值,以识别数据错误或异常情况;所述数据一致性评估模块包括逻辑一致性检测单元和时间一致性检测单元;所述逻辑一致性检测单元通过拓扑关系分析和拓扑错误检查方法分析数据之间的逻辑关系,检测数据是否符合预期的一致性规则,以评估数据的一致性;对于具有时间属性的数据,通过时间一致性检测单元评估数据一致性;所述时间一致性检测单元通过时间序列分析和时间窗口比较方法检测数据在时间上的一致性。

[0042] 作为本发明进一步的技术方案,所述数据纠错引擎包括数据预处理单元、异常识别单元、异常修复单元和数据验证单元;所述数据预处理单元通过删除重复值、处理缺失值以及处理格式错误的方式保证数据的完整性和一致性;所述异常识别单元通过箱线图和孤立森林算法识别测绘数据中的异常值和离群点;所述异常修复单元通过线性插值、多项式插值、回归模型拟合和基于规则的替换方法修复测绘数据中的错误、异常值和离群点;所述

数据验证单元通过数据一致性检查和数据分布分析方法确保修复后的数据的准确性和可靠性。

[0043] 积极有益效果:

[0044] 本发明通过空间分析模块提供了更全面和精细的空间分析能力,揭示了地物的集聚程度和空间相关性;通过统计分析模块全面地了解测绘数据的基本特征和不同变量之间的相关程度;通过智能分析模块对测绘数据进行自动化的分类、预测和优化处理,提高了测绘数据的质量和效率;通过智能评价模块对测绘数据的准确性、完整性和一致性进行评估,并检查和修复测绘数据中的错误和异常值,提高了数据的可靠性和准确性。

附图说明:

[0045] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动性的前提下,还可以根据这些附图获得其他的附图,其中:

[0046] 图1为本发明的流程步骤示意图;

[0047] 图2为本发明的空间分析模块的工作方式原理图;

[0048] 图3为本发明的质量评价模型的工作原理框架图;

[0049] 图4为本发明的数据挖掘方法步骤流程图;

具体实施方式

[0050] 下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0051] 如图1-图4所示,一种测绘数据智能分析评价方法,包括以下步骤:

[0052] 步骤1、通过数据预处理方法对测绘数据进行清洗去噪、校正转换和缺省值填充操作;所述校正转换通过坐标校正和投影转换方法以确保数据的空间准确性和一致性;所述缺省值填充通过插值和统计方法填充缺失的数据值;

[0053] 步骤2、通过图像处理工具从测绘数据中提取特征,并通过特征选择器排除冗余和不相关的特征,以提高后续分析的效率和准确性;

[0054] 步骤3、通过空间分析模块对地理信息进行空间叠加,所述空间分析模块包括邻域分析单元、聚类分析单元和空间关联性分析单元;所述邻域分析单元通过空间索引结构和欧式距离评估地物的邻域密度和距离指标,以确定地物的集聚程度;所述聚类分析单元通过聚类算法识别具有相似空间特征的地物群集;所述空间关联性分析单元通过空间自相关统计方法衡量地物之间的空间相关性;

[0055] 步骤4、通过统计分析模块对测绘数据进行统计描述、统计推断和假设检验处理,以揭示测绘数据的内在规律和特征;所述统计分析模块通过计算均值、方差和百分位数以描述测绘数据特征,并通过相关系数度量不同变量之间的相关程度;

[0056] 步骤5、通过智能分析模块对测绘数据进行分类、预测和优化处理,以提高测绘数

据的质量和效率,

[0057] 步骤6、通过智能评价模块进行数据评价与异常检测;所述智能评价模块包括数据质量评价单元和异常检测单元;所述数据质量评价单元通过数据挖掘方法和质量评价模型对测绘数据的准确性、完整性和一致性进行评估;所述异常检测单元通过数据纠错引擎检查并修复测绘数据中的错误、异常值和离群点,以提高数据的可靠性和准确性;

[0058] 步骤7、通过可视化模块将评价内容和分析后的测绘数据进行展示。

[0059] 在上述实施例中,所述图像处理工具包括特征提取模块、图像分割模块、特征降维模块和后处理模块;所述特征提取模块通过局部二值模式和梯度方向直方图计算和提取特征;所述图像分割模块包括阈值分割单元和边缘分割单元;所述阈值分割单元根据图像亮度或颜色的阈值,通过基于颜色空间的阈值分割方法将图像分为不同的区域;所述边缘分割单元通过索贝尔算子检测图像中物体边缘进行分割,以提取出物体的边界信息;所述特征降维模块包括特征选择单元和特征降维单元;所述特征选择单元通过统计方法和信息论方法评估特征的重要性和相关性,以选取特征;所述特征降维单元通过主成分提取和线性判别分析将高维的特征空间降低到低维空间,以降低特征复杂度;所述后处理模块包括噪声去除单元和目标合并单元;所述噪声去除单元通过滤波器对分类和识别结果中的噪声进行去除,以提高结果的准确性;所述目标合并单元通过连通区域分析方法将相邻的目标进行合并,以形成完整目标区域;所述特征提取模块的输出端与所述图像分割模块的输入端连接;所述图像分割模块的输出端与所述特征降维模块的输入端连接;所述特征降维模块的输出端与所述后处理模块的输入端连接。

[0060] 在具体实施例中,图像处理工具从测绘数据中提取特征的工作方式包括以下步骤:

[0061] S1.对原始测绘数据进行去噪、平滑、增强等预处理操作,以提高图像质量和减少噪声。

[0062] S2.使用计算机视觉和图像分析算法,提取测绘数据中的特征信息。例如,可以使用边缘检测算法寻找道路和建筑物的边界,使用纹理特征提取算法提取地表纹理信息,使用形状分析算法提取不同对象的形状等。

[0063] S3.将测绘数据图像分割成不同的区域或对象。通过采用聚类、阈值分割、区域生长等算法,将图像中相似的像素或区域归类为同一对象或区域。

[0064] S4.在特征空间中进行降维操作,以减少特征的数量同时保留有意义的信息。常用的方法包括主成分分析(PCA)、线性判别分析(LDA)等。

[0065] S5.对提取的特征图像进行形态学操作,如开闭运算、腐蚀、膨胀等,以消除小的噪声和非重要结构。同时对特征进行组合和整合,生成更复杂和全面的特征描述。

[0066] 图像处理工具利用了测绘数据中的视觉特征,并通过适当的算法和模型进行分析和处理。通过提取关键特征和区域,以及降低数据的维度,可以从测绘数据中获取更有意义和有益的信息。

[0067] 在具体实施中,通过图像处理工具,可以从测绘数据中提取出关键的地理特征,如道路、建筑物、河流等,使其更易于分析和解释。通过特征提取和图像分割,可以将测绘数据转化为可量化和分类的形式,使其更适合于后续的数据分析和应用。另外,从测绘数据中提取的特征信息可以为决策制定提供重要的依据和参考,帮助进行城市规划、资源管理等方

面的决策。通过图像处理工具,可以将测绘数据转化为图像或可视化结果,使其更加直观和易于理解,有助于数据交流和共享,具体实施中,图像处理工具性能评估表如表1所示:

[0068] 表1图像处理工具性能评估表

样本编号	图像分辨率	噪声水平	对比度	锐化程度	颜色偏移	模糊程度	分类结果	
	1	512x512	0.05	0.8	0.7	0.1	0.2	高质量
	2	1024x768	0.1	0.6	0.9	0.3	0.4	高质量
[0069]	3	256x256	0.2	0.4	0.6	0.2	0.6	中等质量
	4	1920x1080	0.05	0.9	0.8	0.4	0.1	高质量
	5	640x480	0.3	0.3	0.4	0.1	0.8	低质量
	6	1280x720	0.2	0.7	0.5	0.2	0.4	中等质量
样本编号	图像分辨率	噪声水平	对比度	锐化程度	颜色偏移	模糊程度	分类结果	
	7	800x600	0.1	0.5	0.6	0.3	0.3	中等质量
[0070]	8	2560x1440	0.05	0.9	0.7	0.1	0.1	高质量
	9	320x240	0.3	0.2	0.3	0.2	0.7	低质量
	10	3840x2160	0.1	0.8	0.6	0.4	0.2	高质量

[0071] 数据表格1中,图像分辨率表示图像的宽度和高度,噪声水平表示图像中的干扰噪声强度,对比度表示图像中不同亮度级别之间的差异程度,锐化程度表示图像边缘的清晰度程度,颜色偏移表示图像中颜色的偏移程度,模糊程度表示图像的模糊程度。分类结果表示根据这些特征对图像质量进行的分类。

[0072] 在上述实施例中,所述邻域分析单元通过邻域密度指标对地物的邻域密度进行分析,并通过R树对高维数据和多边形对象进行数据化分析;其中,领域分析的公式表达式为:

[0073]
$$K = \log \sum e^y \sqrt{y+q} - F \frac{-q \pm \sqrt{F^2 - 4r}}{2e^y} + r \quad (1)$$

[0074] 在公式(1)中,K表示邻域密度,y表示邻域内点的数量,r表示邻域半径,q表示邻域内线段的长度,F表示邻域内面积;

[0075] 所述聚类分析单元通过迭代实现最小化总体误差;迭代步骤为:

[0076] (S1) 初始化聚类中心;

[0077] (S2) 计算样本与聚类中心的距离,将样本划分到最近的类别;

[0078] (S3) 更新聚类中心为每个类别的均值;

[0079] (S4) 重复步骤(S2)和(S3),直到满足停止条件;

[0080] 所述空间关联性分析单元通过莫兰指数衡量空间数据分布的集聚程度,莫兰指数的公式表达式为:

$$[0081] \quad I = \frac{W}{N} \prod \sec^{-1} W + \log \sqrt{i^2 + j^2} - \cos N (j - i)^2 \quad (2)$$

[0082] 在公式(2)中,N表示样本点数量,W表示空间邻接矩阵,i表示空间距离权重参数,j表示参数对应的属性值;另外,空间关联性分析单元通过吉尔利指数衡量空间数据的离散程度,吉尔利指数的公式表达式为:

$$[0083] \quad C = \frac{W}{N} \sum \sec^{-1} (j - i) + \frac{\sqrt{i^2 - 4j}}{M} - W \log \frac{\pi}{2} \sqrt{i^2 + j^2} - N \cos N^2 \quad (3)$$

[0084] 在公式(3)中,N表示样本点数量,W表示空间邻接矩阵,i表示空间距离权重参数,j表示参数对应的属性值,M表示所有样本点的属性值均值;吉尔利指数的取值范围为[0,2],当值接近0时,表示正相关性,即相似值与相似值相邻;当值接近2时,表示负相关性,即相似值与不相似值相邻;当值接近1时,表示无空间自相关性。

[0085] 在具体实施例中,邻域分析单元用于计算测绘数据中每个地理信息对象周围的邻域特征。例如,可以确定一个地理信息对象周围的其他地理信息对象的数量、密度或距离等。常用的邻域分析方法包括缓冲区分析、近邻分析等。

[0086] 聚类分析单元用于将测绘数据中的地理信息对象根据其空间位置或属性进行聚类。这有助于发现具有相似特征的地理信息群体和模式。常用的聚类分析方法包括K均值聚类、DBSCAN等。

[0087] 空间关联性分析单元用于评估测绘数据中地理信息对象之间的空间关系。它可以识别对象之间的相互作用、依赖或群集趋势等。常用的空间关联性分析方法包括空间自相关、热点分析等。

[0088] 在具体实施例中,通过邻域分析单元,可以了解地理信息对象周围的环境特征,如密度、数量、距离等。这对于城市规划、土地利用分析等具有重要意义。

[0089] 通过聚类分析单元,可以发现具有相似特征的地理信息群体,识别出地理模式和趋势,为决策制定提供依据。例如,可以将城市区域划分为不同的功能区域,如商业区、住宅区等。

[0090] 通过空间关联性分析单元,可以评估地理信息对象之间的空间关系,揭示地理现象的相互作用和依赖关系。这有助于理解城市发展、资源分配等方面的问题。

[0091] 基于空间分析模块的结果,可以为决策制定提供支持。通过了解地理信息对象的空间分布和特征,可以更好地进行城市规划、资源管理、应急响应等决策。

[0092] 具体实施中,不同地理信息对象的空间特征、聚类结果和空间关联性信息的表示如表2所示:

[0093] 表2空间分析测试数据表

[0094]	地理信息对象	X 坐标	Y 坐标	邻域特征	聚类结果	空间关联性
	地物 A	10.0	15.0	5	集群 1	正相关
	地物 B	12.0	20.0	3	集群 2	负相关
[0095]	地物 C	8.0	18.0	4	集群 1	正相关
	地物 D	25.0	30.0	2	集群 3	中性相关
	地物 E	30.0	35.0	6	集群 3	正相关

[0096] 表格2中,每一行代表一个地理信息对象,包括其位置坐标(X,Y)、邻域特征、聚类结果和空间关联性,空间分析模块通过对测绘数据进行邻域分析、聚类分析和空间关联性分析,可以提取地理信息的空间特征、发现地理模式和趋势,揭示地理现象之间的关系,并为决策制定提供支持。这些有益效果可以帮助解决城市规划、资源管理、应急响应等领域的问题。

[0097] 在上述实施例中,所述统计分析模块包括统计描述单元、统计推断单元和假设检验处理单元;所述统计描述单元通过描述性统计和极差分析描述和总结测绘数据的特征和分布情况;所述描述性统计通过数据的均值、中位数、众数和标准差获得数据中心趋势和离散程度,以描述数据的集中程度和分散程度;所述极差分析通过数据最大值和最小值之间的差异衡量数据的变异程度;所述统计推断单元基于从测绘数据中获取的样本数据,通过参数估计和置信区间估计方法推断总体的性质和特征;所述假设检验处理单元通过集计分析、相关分析和回归分析验证或否定在测绘数据分析中提出的假设;所述集计分析通过交叉表分析和频数分布分析方法了解数据要素之间的关系;所述相关分析通过皮尔逊相关系数判断数据之间的线性关系强度和方向;所述回归分析通过线性回归函数关系模型预测未知数据。

[0098] 在具体实施例中,统计分析模块通过统计描述单元对测绘数据进行整理、归纳和总结,提供数据的基本特征和描述性统计信息,例如平均值、标准差、最大值、最小值等。通过统计推断单元基于样本数据,利用统计学原理和方法进行参数估计和假设检验。统计推断单元可以根据样本数据推断总体数据的特征,例如总体均值的置信区间、相关性的显著性检验等。通过假设检验处理单元验证关于总体特征的假设,通过设定显著性水平并进行统计检验,判断某种假设是否成立或拒绝。常用的假设检验方法包括t检验、方差分析、卡方检验等。

[0099] 具体实施例中,通过统计分析模块,可以提供客观的数据描述和分析结果,使决策者能够更好地理解测绘数据的特征和规律。同时,可以为决策提供科学依据,帮助评估测绘

数据质量、准确性和一致性,进而指导后续工作和决策制定。另外,可以发现数据中存在的异常值、异常情况或趋势,帮助及早发现问题并采取相应的措施进行纠正。其次,在数据比较、差异分析和相关性分析中起到重要作用,揭示数据之间的关系和影响因素,有助于深入理解测绘数据所涉及的问题领域。同时,通过提供数据的置信区间和显著性检验结果,使决策者能够更准确地评估结果的可靠性和有效性。

[0100] 通过统计分析模块,测绘数据智能分析评价方法可以更全面、准确地对测绘数据进行分析和评估,为决策提供科学依据,促进测绘工作的优化和改进。

[0101] 在上述实施例中,所述智能分析模块包括特征工程单元和模型训练评估单元;所述特征工程单元通过特征选择和特征转换的方式对测绘数据进行预处理,以提取特征;所述特征选择通过信息增益衡量特征对分类问题的贡献度;并通过相关系数计算特征与目标变量间的相关性,以提取目标特征;所述特征转换通过主成分提取和线性判别分析将原始特征转换为新的特征表示,以提高模型性能;所述主成分提取通过将多维数据降维为低维度表示,以消除冗余信息,提取关键特征;所述线性判别分析通过将高维数据投影到低维空间,并保留区分不同类别的信息,以实现分类问题的优化和提高;所述模型训练评估单元通过随机森林算法训练模型,并通过评估指标对模型的性能进行评估,所述评估指标包括准确率、召回率和调和平均数。

[0102] 在具体实施例中,智能分析模块通过特征工程单元对原始的测绘数据进行处理和转换,提取出与评价指标相关的有效特征。这包括数据清洗、特征选择、降维等技术,以确保模型输入的特征具有高度的表达能力和区分度。通过特征工程,可以有效减少噪声的影响,提高模型的鲁棒性和准确性。另外,通过模型训练评估单元使用经过特征工程处理后的数据作为输入,利用机器学习算法构建预测模型。该模型可以对测绘数据进行分类、预测和优化处理。针对不同的评价指标,选择合适的算法模型(如决策树、支持向量机、神经网络等)进行训练,并通过交叉验证、指标评估等方法对模型进行评估和调优,以获得更好的预测效果。

[0103] 具体实施中,智能分析模块能够自动处理和分析海量的测绘数据,代替传统的人工分析方法,大大提高了工作效率。通过模型训练和评估,智能分析模块能够基于历史数据和特征来进行准确的预测,帮助用户更好地了解和评估测绘数据中的各种情况和趋势。另外,智能分析模块可以根据特征分析结果,采取针对性的优化处理方法,提高测绘数据的质量和准确性。基于智能分析模块的结果和预测,决策者可以做出更明智的决策,优化资源分配和规划。通过智能分析模块对测绘数据进行分类、预测和优化处理,可以实现自动化、高效率和高准确性的数据分析,为测绘工作和决策提供有力支持。同时,智能分析模块能够自动执行各种数据处理和分析任务,从数据清洗、特征提取到模型训练和评估,无需人工干预,大大节省了人力资源和时间成本。其次,智能分析模块通过不断接收和处理新的测绘数据,模型可以进行迭代和更新,进一步优化分析和预测结果。同时,智能分析模块可以同时考虑多个评价指标,并对其进行综合评估。通过对不同指标之间的权重进行设定和调整,可以根据实际需求对测绘数据进行全面的评价和分析,具体实施中,数据智能分析表格如表3所示:

[0104] 表3数据智能分析表

序号	输入数据	特征工程结果	模型预测结果	优化处理结果	期望结果	是否通过	
[0105]	1	测绘数据集 1	特征 1: 0.85, 特征 2: 0.72	类别 A	优化结果 A	类别 A, 优化结果 A	是
	2	测绘数据集 2	特征 1: 0.45, 特征 2: 0.61	类别 B	优化结果 B	类别 B, 优化结果 B	是
	3	测绘数据集 3	特征 1: 0.76, 特征 2: 0.83	类别 A	优化结果 A	类别 A, 优化结果 A	是
	4	测绘数据集 4	特征 1: 0.62, 特征 2: 0.57	类别 C	优化结果 C	类别 C, 优化结果 C	是

[0106] 在数据表格3中,序号用于对不同的测试案例进行编号,输入数据作为输入到智能分析模块的测绘数据集,特征工程结果经过特征工程单元处理后得到的测绘数据特征结果,模型预测结果使用模型训练评估单元对特征工程结果进行预测的输出结果,优化处理结果基于模型预测结果进行的优化处理或建议,期望结果根据专家知识或预先设定的指标,针对每个测试案例期望得到的结果,是否通过记录该测试案例是否通过,即智能分析模块的输出结果是否符合预期。

[0107] 在上述实施例中,所述可视化模块包括可视化展示单元和数据整合单元;所述数据整合单元通过数据集成分析方法将空间分析模块、统计分析模块和智能分析模块的分析结果进行整合,以得出综合评价结果,最终通过可视化展示单元进行展示;所述可视化展示单元通过地理信息系统和数据可视化工具,将分析结果以地图或图表形式展示,以帮助用户理解和使用测绘数据;并通过报告生成器根据分析结果生成文档报告。

[0108] 在具体实施例中,可视化模块会从智能分析模块获取评价结果数据。这些数据可能包括不同地点、不同时间或不同指标的评价结果。可视化模块将对这些数据进行整合和准备,以便后续的可视化展示。同时,可视化模块将准备好的数据转换为可视化图表、图形或地图等形式。例如,可以使用柱状图、折线图、热力图、散点图等来展示测绘数据的评价情况和趋势。根据具体需求,可以选择不同的可视化方式来呈现数据。

[0109] 在可视化展示单元中,用户通常可以与图表或地图进行交互,并进行数据的筛选和调整。用户可以选择特定的地区、时间范围或指标来观察不同的评价结果。通过数据交互和筛选功能,用户可以更深入地探索和分析测绘数据。

[0110] 在可视化展示过程中,可视化模块还应提供结果解释和洞察的功能。这包括对评价结果的解释、异常数据的识别、趋势分析和量化指标的展示等。通过提供解释和洞察,可视化模块可以帮助用户理解数据背后的意义和价值,并支持决策过程。

[0111] 在具体实施中,可视化图表和地图能够直观地展示测绘数据的评价结果,使用户

能够快速理解和分析数据。通过图形化的方式,复杂的数据可以更容易地被人理解和解释。同时,可视化模块能够帮助用户发现数据中的隐藏模式和趋势。通过观察和比较不同的图表或地图,用户可以识别出测绘数据中的规律性变化或异常情况,进而探索数据背后的原因和因素。另外,可视化模块提供的结果解释和洞察功能可以为用户的决策提供支持。通过深入理解数据评价结果,用户可以做出更准确和明智的决策,从而提高工作效率和决策质量。其次,可视化模块生成的图表或地图可以作为沟通和共享的工具。用户可以将可视化结果以图片、报告或在线平台的形式分享给团队成员或相关利益相关者,促进交流和合作。

[0112] 在上述实施例中,所述数据挖掘方法的工作方式步骤为:

[0113] (S1) 通过测绘数据采集设备获取测绘数据;

[0114] (S2) 通过数据清洗工具进行去除重复值、处理缺失值和处理异常值操作,以确保数据的准确性和完整性;

[0115] (S3) 通过特征选择算法从原始数据中选择与评估目标相关的特征;所述特征选择算法包括信息增益、相关系数提取和主成分提取操作;

[0116] (S4) 通过特征转换方法将选择的特征进行转换,以便于后续分析和建模;

[0117] (S5) 通过关联规则挖掘算法,构建评估模型;

[0118] (S6) 使用已标注的数据集对构建的模型进行评估,比较模型预测结果与实际情况的差异,并计算评估指标;

[0119] (S7) 根据评估结果,对测绘数据的准确性、完整性和一致性进行解释和分析,发现数据的潜在问题和规律,并提出改进建议。

[0120] 在具体实施例中,数据挖掘方法通过数据预处理操作对原始测绘数据进行清洗、去噪和转换等处理,以消除数据中的异常值、缺失值和冗余信息,并将数据转化为适合挖掘的形式。其次,基于预处理后的数据,通过特征提取和选择操作提取与测绘数据质量相关的特征,如数据的精度、完整度、一致性等,同时根据特征的重要性进行选择,排除对评价结果影响较小的特征。另外,基于挖掘的目标和评价标准,通过构建评价模型,选择合适的数据挖掘算法,如决策树、神经网络、支持向量机等,以训练出一个高效的模型。利用标记好的数据对模型进行训练和优化,调整和优化参数以提高评价模型的准确性和泛化能力。最后,应用模型评价未标记的测绘数据,获取评价结果,并进行验证和评估。根据评估结果,可对测绘数据质量进行准确性、完整性和一致性等方面的评估。

[0121] 这些操作可以帮助提高测绘数据评估的效率和准确性,自动化地对大量数据进行处理和分析,节省人力资源和时间成本;同时,基于数据挖掘方法的评估模型能够全面考虑多个质量指标,得出综合评价结果,更准确地判断测绘数据的质量水平;此外,数据挖掘方法可以帮助发现数据中存在的错误、缺失和不一致情况,提供针对性的改进建议,促进测绘数据的质量提升;同时,数据挖掘方法可以揭示数据之间的关联性和规律性,为后续的数据分析和决策提供有用的信息;最后,通过不断优化和更新评价模型,可以适应不同类型的测绘数据和评估需求,提高测绘数据评价的可靠性和适用性。

[0122] 在上述实施例中,所述质量评价模型包括数据准确性评估模块、数据完整性评估模块和数据一致性评估模块;所述数据准确性评估模块包括空间一致性评估单元和属性一致性评估单元;所述空间一致性评估单元通过重叠度计算和误差椭圆生成方法比较同一区域内不同数据源采集的相同地理要素,检测数据之间的几何一致性;所述属性一致性评估

单元通过属性值比较和相似度计算方法比较同一地理要素在不同数据源中的属性值,以评估数据的属性准确性;所述数据完整性评估模块包括缺失值检测单元和异常值检测单元;所述缺失值检测单元通过缺失值比例统计和缺失机制分析的方式分析数据中的缺失情况,识别缺失数据的位置和类型;以评估数据完整性;所述异常值检测单元通过离群点检测和异常值分析方法检测数据中的异常值,以识别数据错误或异常情况;所述数据一致性评估模块包括逻辑一致性检测单元和时间一致性检测单元;所述逻辑一致性检测单元通过拓扑关系分析和拓扑错误检查方法分析数据之间的逻辑关系,检测数据是否符合预期的一致性规则,以评估数据的一致性;对于具有时间属性的数据,通过时间一致性检测单元评估数据一致性;所述时间一致性检测单元通过时间序列分析和时间窗口比较方法检测数据在时间上的一致性。

[0123] 在具体实施例中,质量评价模型通过数据准确性评估模块收集测绘数据的参考数据或真实测量数据作为验证标准,并将测绘数据与验证标准进行比较,计算测绘数据与实际数据之间的差异。另外,根据差异的大小和特征,确定准确性评估指标,如均方根误差(RMSE)或偏差百分比等。最后,应用合适的统计分析或机器学习方法对准确性指标进行分析,得出准确性评价结果。

[0124] 数据准确性评估模块通过与真实数据的比对,提供了测绘数据的准确性度量,帮助用户了解数据的质量水平。另外,可以发现数据中存在的精度问题,帮助用户识别潜在的错误和误差。基于准确性评价结果,可以采取相应的措施来改善数据质量,提高测绘数据的可靠性。

[0125] 具体实施中,数据完整性评估模块的工作方式如下所示:

[0126] a. 确定衡量数据完整性的指标,如缺失率、覆盖范围等。

[0127] b. 对测绘数据进行检查,识别缺失或不完整的数据部分。

[0128] c. 基于指标和检查结果,计算数据的完整性得分或比例。

[0129] d. 进行数据统计和分析,得出完整性评价结果。

[0130] 数据完整性评估模块通过度量数据的完整性,帮助用户了解数据中存在的缺失情况,提高对数据的理解。另外,可以发现数据集中缺失的部分,以便采取措施收集丢失的数据,增强数据的完整性。其次,根据完整性评价结果,可以选择性地使用或排除不完整的数据,提高后续分析和应用的可靠性。

[0131] 数据一致性评估模块的工作方式如下所示:

[0132] a. 确定数据一致性的评价指标,如逻辑一致性、拓扑一致性等。

[0133] b. 对测绘数据进行一致性检查,识别可能存在的矛盾、冲突或错误。

[0134] c. 基于一致性指标和检查结果,计算数据的一致性得分或比例。

[0135] d. 对一致性问题进行分析和处理,得出一致性评价结果。

[0136] 数据一致性评估模块通过评估数据的一致性,帮助用户识别潜在的冲突和错误,提高对数据的可信度。同时,通过数据一致性评估模块可以发现数据中存在的逻辑或拓扑上的不一致情况,以便进行进一步的数据校正和验证。其次,基于一致性评价结果,可以提供修复方法或建议,保证数据的一致性,增强数据的应用价值。

[0137] 总体而言,这种测绘数据智能分析评价方法通过质量评价模型的工作方式原理步骤,对测绘数据的准确性、完整性和一致性进行全面评估。这样的方法可以帮助用户全面了

解测绘数据的质量状况,发现潜在的问题和改进空间,并提供相应的措施和建议,以提高测绘数据的质量和可靠性。最终,使用者能够更加有效地利用测绘数据进行分析 and 决策;具体实施中,测绘数据质量评估结果如表4所示:

[0138] 表4测绘数据质量评估表

	数据指标	数据样本 1	数据样本 2	数据样本 3	数据样本 4
[0139]	准确性	0.452	0.618	0.512	0.675
	完整性	0.092	0.071	0.085	0.062
	数据指标	数据样本 1	数据样本 2	数据样本 3	数据样本 4
[0140]	一致性	4	2	3	5
	可信度	0.846	0.921	0.793	0.908
	时效性	8	10	9	7

[0141] 在上述实施例中,所述数据纠错引擎包括数据预处理单元、异常识别单元、异常修复单元和数据验证单元;所述数据预处理单元通过删除重复值、处理缺失值以及处理格式错误的方式保证数据的完整性和一致性;所述异常识别单元通过箱线图和孤立森林算法识别测绘数据中的异常值和离群点;所述异常修复单元通过线性插值、多项式插值、回归模型拟合和基于规则的替换方法修复测绘数据中的错误、异常值和离群点;所述数据验证单元通过数据一致性检查和数据分布分析方法确保修复后的数据的准确性和可靠性。

[0142] 在具体实施例中,数据纠错引擎通过数据预处理单元用于对测绘数据进行初步清洗和准备工作。它可以包括去除重复数据、填充缺失值、处理异常数据等操作。数据预处理的目的是为后续的异常识别和修复提供高质量的数据基础。通过异常识别单元利用算法和模型来检测数据中的异常情况。它可以通过统计分析、机器学习或其他技术手段来识别与预期数据分布不一致的数据点。异常识别的目的是找出潜在的数据错误、缺陷或异常情况。通过异常修复单元根据异常识别结果,对检测到的异常数据进行修正或纠正。修复方法可以根据具体的异常类型采用插值、外推、替换或其他合适的算法进行修复。异常修复的目的是恢复数据的准确性和完整性,使其符合预期的数据规范。通过数据验证单元对修复后的数据进行再次验证,以确保异常数据已经得到了正确的修复。它可以通过比较修复后的数据与原始数据、检查数据的一致性和逻辑关系等方式进行验证,以确保修复结果的准确性和可靠性。

[0143] 在具体实施中,通过自动化的纠错过程,可以有效减少数据错误和异常情况,提高测绘数据的准确性、完整性和一致性,从而提高整体数据质量。另外,数据纠错引擎能够帮助发现和修复异常情况,增强数据的可信度和可靠性,使其更具有参考和应用价值。其次,使用数据纠错引擎可以自动化处理数据纠错过程,节省人力资源和时间成本,提高工作效率。

率和生产效率。经过数据纠错引擎的处理,清洗和修复后的数据更加准确和可靠,可以为后续的数据分析、模型训练和决策制定提供更可靠的基础。

[0144] 虽然以上描述了本发明的具体实施方式,但是本领域的技术人员应当理解,这些具体实施方式仅是举例说明,本领域的技术人员在不脱离本发明的原理和实质的情况下,可以对上述方法和系统的细节进行各种省略、替换和改变。例如,合并上述方法步骤,从而按照实质相同的方法执行实质相同的功能以实现实质相同的结果则属于本发明的范围。因此,本发明的范围仅由所附权利要求书限定。

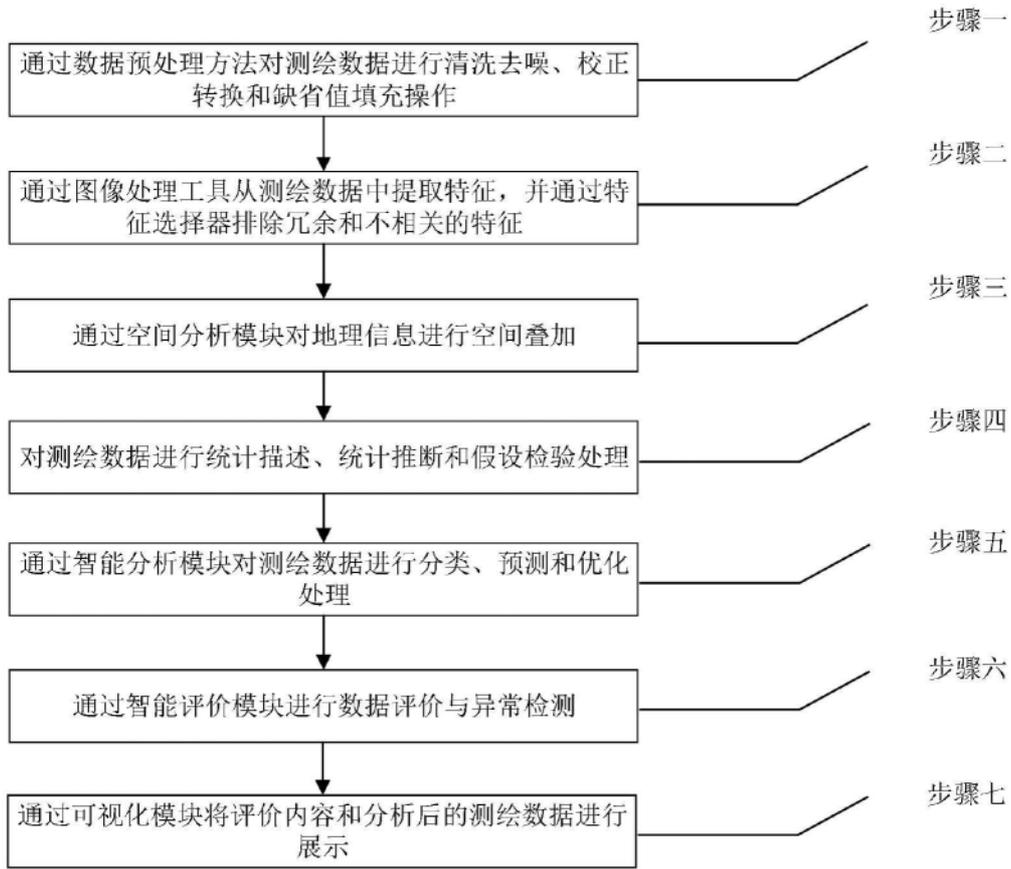


图1

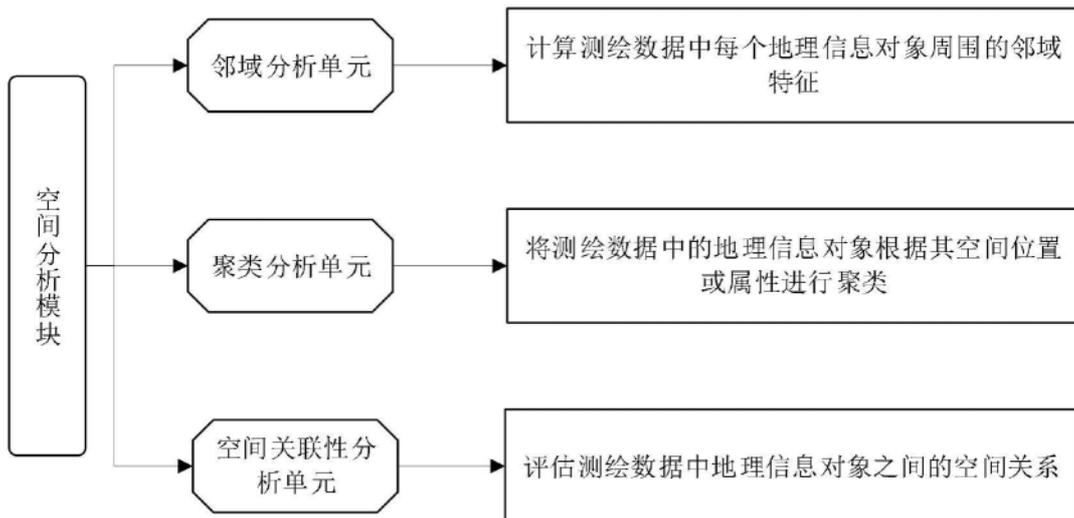


图2

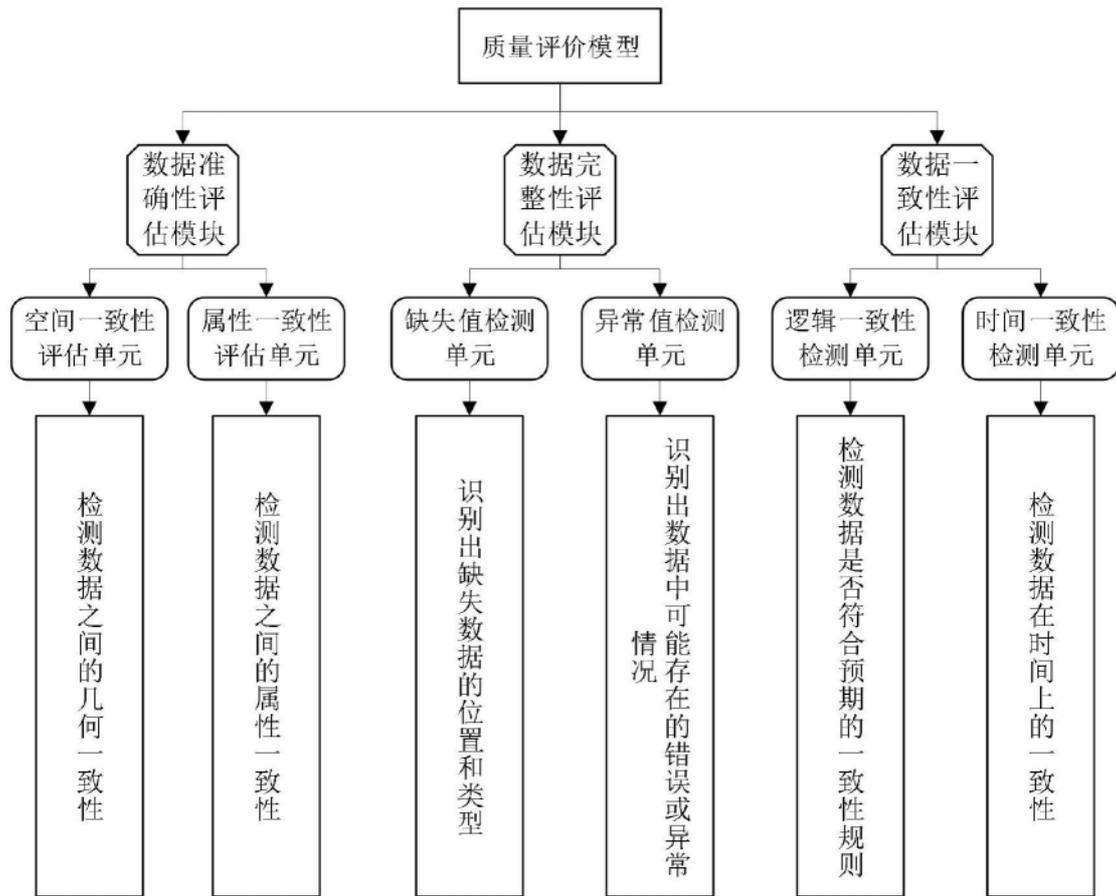


图3

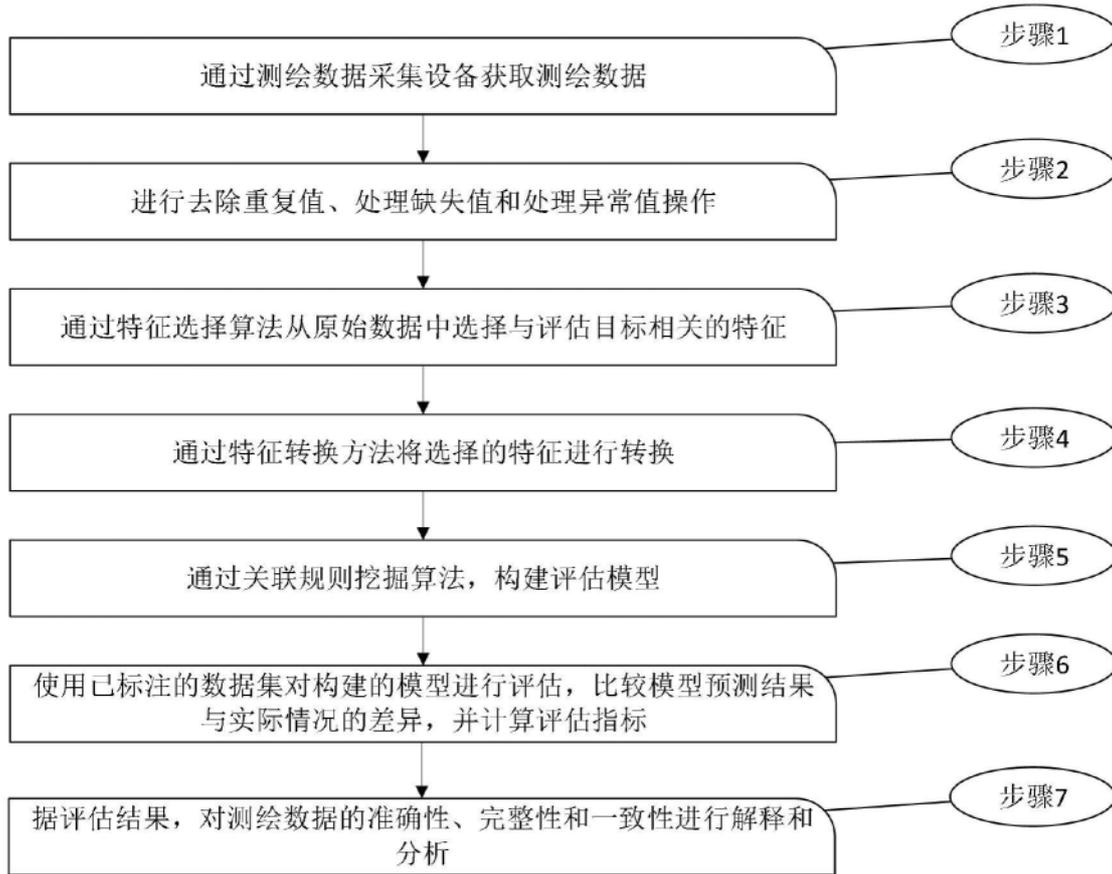


图4