

AI 与智慧水务

刘彬, 苏晓, 田宝义, 等. 供水管网 GIS 数据质量提升案例分析 [J]. 净水技术, 2025, 44(11): 165-173, 194.

LIU B, SU X, TIAN B Y, et al. Case analysis on GIS data quality upgrading for water supply pipelines network [J]. Water Purification Technology, 2025, 44(11): 165-173, 194.

供水管网 GIS 数据质量提升案例分析

刘彬¹, 苏晓¹, 田宝义¹, 张巧玉^{2,*}, 彭森²

(1. 天津水务集团有限公司, 天津 300202; 2. 天津大学环境科学与工程学院, 天津 300350)

摘要 【目的】城市供水管网地理信息系统(GIS)数据是智慧水务建设的关键基础,其数据质量直接影响管网风险评估、故障预警及运维效率,当前GIS数据质量管理仍面临系统性不足的挑战,缺乏从数据预处理、错误修正到质量评估的完整技术链条,难以实现复杂管网多源异构数据的全流程质量提升。【方法】本文提出一种融合拓扑结构与属性特征的供水管网GIS数据质量提升框架:首先,基于深度优先搜索(depth-first search, DFS)和Dijkstra算法构建拓扑连通性诊断及自动修正方法,实现管网拓扑错误的高效检查与修正;其次,结合K近邻(k-nearest neighbors, KNN)算法与反距离加权(inverse distance weighting, IDW)插值算法,实现管网多尺度属性数据的逻辑检查与缺失填补;最后,建立综合性质量评估指标,量化验证数据提升效果。【结果】案例分析表明,数据质量提升框架的实施将数据质量评估指数平均值由处理前的92.36%提升至99.92%,表明算法显著提升了供水管网数据的整体质量。同时,相较于传统GIS软件处理方式,该框架减少了人工检查的工作量,提高了数据处理效率与修正准确率,验证了框架的高效性与可靠性。【结论】本文为供水管网GIS数据质量提升提出了系统性的方法,显著提升了数据质量与处理效率,为智慧水务系统的高效运行提供了重要支撑。

关键词 供水管网 地理信息系统(GIS)数据质量提升 拓扑结构修正 属性数据填补 深度优先搜索算法 Dijkstra算法

中图分类号: TU991 **文献标志码:** A **文章编号:** 1009-0177(2025)11-0165-10

DOI: 10.15890/j.cnki.jsjs.2025.11.019

Case Analysis on GIS Data Quality Upgrading for Water Supply Pipelines Network

LIU Bin¹, SU Xiao¹, TIAN Baoyi¹, ZHANG Qiaoyu^{2,*}, PENG Sen²

(1. Tianjin Water Group Co., Ltd., Tianjin 300202, China;

2. School of Environmental Science and Engineering, Tianjin University, Tianjin 300350, China)

Abstract [Objective] Urban water supply network geographic information system (GIS) data serves as a critical foundation for smart water management systems. Their quality directly influences pipe network risk assessment, fault warning, and operational efficiency. Current GIS data quality management still faces systemic challenges, particularly the lack of an integrated technical chain encompassing data preprocessing, error correction, and quality evaluation, which hinders comprehensive quality improvement of multi-source heterogeneous data in complex pipe networks. [Methods] A GIS data quality improvement framework integrating topological structure and attribute features was proposed in this study. First, a topological connectivity diagnosis and automated correction method was constructed using depth-first search (DFS) and Dijkstra's algorithm to efficiently detect and rectify topological errors. Second, logical verification and missing data imputation for multi-scale attribute data were achieved by combining k-nearest neighbors (KNN) with inverse distance weighting (IDW) interpolation algorithms. Finally, comprehensive quality evaluation metrics were established to quantitatively validate data improvement outcomes. [Results] Case analysis indicated that the implementation of the data quality

[收稿日期] 2025-02-23

[基金项目] 天津水务集团有限公司科技项目(2023KY-04)

[作者简介] 刘彬(1980—),男,高级工程师,研究方向为供水管网系统建设、运行管理与风险评估,E-mail:13820071311@163.com。

[通信作者] 张巧玉(2000—),女,硕士研究生,研究方向为供水管网系统地理信息数据分析与漏损控制,E-mail:2311662989@qq.com。

improvement framework increased the average value of the data quality assessment index from 92.36% before processing to 99.92%, demonstrating that the algorithm significantly improved the overall quality of the water supply network data. Additionally, compared to traditional GIS software processing method, this framework reduced the workload of manual checks and improved data processing efficiency and correction accuracy, validating the framework's effectiveness and reliability. [Conclusion] This paper puts forward a systematic method for enhancing the quality of GIS data in water supply networks, significantly improving data quality and processing efficiency, thus providing crucial support for the efficient operation of smart water management systems.

Keywords water supply pipelines network geographic information system (GIS) data quality upgrading topological structure correction attribute data imputation depth-first search(DFS) algorithm Dijkstra's algorithm

供水管网作为城市基础设施的重要组成部分,对保障居民生活质量、促进经济发展具有深远意义。随着城市化进程加速,人口的增长和城市的扩展使得供水需求不断增加,要求供水管网不仅要不断扩展,还要及时进行更新和维护。这对供水管网数据的质量管理提出了更高的要求,尤其是地理信息系统(GIS)技术的应用,使得供水管网的管理能够更加精准、高效,从而提高供水服务的安全性、可靠性和持续性。

GIS 技术凭借其在空间数据管理和分析方面的优势,已成为供水管网管理中不可或缺的工具。供水管网的空间位置、管道属性以及各类拓扑关系通过 GIS 得以全面整合,为供水系统的优化与决策提供了强有力的技术支撑^[1-2],并在智慧水务的实践中发挥重要作用^[3]。然而,供水管网 GIS 数据常存在诸多质量问题,包括数据不完整、误差、冗余以及拓扑错误,这些问题导致数据的一致性和准确性难以保障,对管网运维造成负面影响。当前 GIS 数据在供水管网风险评估^[4]、风险预测^[5-6]、故障分析^[7-9]及实时监测平台^[10]等多个应用领域中得到广泛应用,但关于提高 GIS 数据质量的研究仍需要进行深入探讨。当前国内一些供水企业已经开始从企业标准层面重视管网 GIS 数据质量问题^[11],但必须发展能提升数据质量的技术,才能从根本上解决问题。因此,如何高效处理 GIS 数据并提高其质量,成为供水管网维护和管理中的关键技术之一。

近年来,国内外学者开始重视管网 GIS 数据质量问题,开展了相关的 GIS 数据处理与分析研究。例如,对于缺失值的填补,多数采用人工或统计学方法,如互补填补法^[12]、规则填补法^[13]或中值填补方法^[14]。然而,这些传统方法在实际应用中往往未能充分考虑管网数据的空间拓扑特性或逻辑关系,导致填补结果与实际情况存在偏差,影响后续分析的

准确性。此外,对于拓扑结构的检查,研究者常用管网建模软件计算出错误位置,并通过人工进行修复和校核^[15],或放到地理信息系统中人工调整^[16]。虽然一些管网建模软件能够识别拓扑问题,但在处理大规模供水管网数据时效率较低,且在与其他 GIS 属性数据融合时操作繁琐。因此,如何在处理大规模供水管网 GIS 数据时提高效率,减少人工操作,成为一个亟待解决的难题。为解决这些问题,当前一些最短路径算法,如广度优先搜索算法和 Dijkstra 算法^[17],在供水管网的连通性检查中得到了应用,这为拓扑结构的检查和修复提供了有效的思路。近年来,GIS 系统在供水管网的拓扑管理^[18]和与水力模型的集成^[19]方面取得了较大进展,但尚未形成从数据预处理到数据质量评估一套完整的处理 GIS 数据的质量提升方法。

基于以上背景,本文构建了供水管网 GIS 数据质量提升的技术框架,对管网数据中的拓扑结构和属性数据进行自动化检查和修正,从而提高 GIS 数据的完整性、一致性和准确性。此外,本文还提出了 GIS 数据质量评估方法,对数据质量进行定量化评估,为后续的数据处理提供参考。上述技术方法能够为供水管网的 GIS 数据管理提供更加可靠的技术支持,从而提高供水管网系统运维整体效率与安全性。

1 供水管网 GIS 数据质量提升技术方法构建

1.1 管网 GIS 数据问题分析

供水管网作为城市基础设施的核心组成部分,涵盖了复杂的设施系统,每个设施不仅在物理空间上有明确的几何分布,同时还包含大量的属性信息。这些 GIS 数据可能来自不同部门、时间,且往往缺乏统一的数据标准和规范,造成数据质量问题。

供水管网 GIS 数据中常见的质量问题主要包括

以下两方面。(1)拓扑结构错误:数据采集或编辑过程中的错误,可能导致节点重复、管道重叠,或者出现孤立节点和孤立管段,从而影响管网的连通性分析。(2)属性信息错误:供水管网中管道的关键属性,如管道的尺寸、材质、年限等,可能因测量误差、输入错误或数据采集不全而出现错误或缺失。这些属性数据的不准确性会直接影响管网的性能评估和故障诊断。

1.2 方法介绍

针对上述问题,本文提出了以下技术框架进行数据质量提升分析与处理,主要包括3个模块,总体框架如图1所示。拓扑检查与修正模块:通过检查节点坐标和管段无序对,识别并删除重复节点和管段;利用基于栈的深度优先搜索(depth-first search, DFS)算法进行连通性检查;结合Dijkstra算法与节

点物理属性加权计算找到与连通主网络最优连接路径,修复拓扑断裂。属性检查与修正模块:通过设置属性数据的合理范围及局部属性判别公式检查全局与局部异常值,同时完成空值检查;利用反距离权重(inverse distance weighting, IDW)插值法及K近邻(k-nearest neighbors, KNN)算法分别对有问题的数据属性和分类属性进行修正。数据质量评估模块:结合管网设计图等资料对算法识别的数据问题进行人工复检,核实修正情况,去除误判的“修正”数据,得到算法真正修正的GIS数据;针对拓扑数据的重复与断裂问题,设置拓扑独特指数和拓扑连通指数进行质量评估;针对属性数据异常与缺失问题,设置数据逻辑指数和数据完整指数进行质量评估;最后引入管网实际案例,对比修正前后GIS数据质量评估指数,分析算法修正效果。

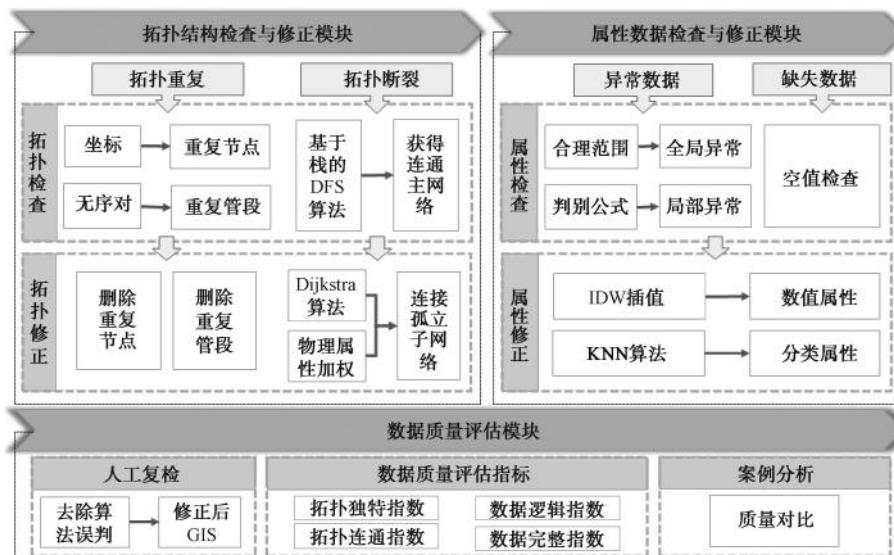


图1 整体技术框架
Fig. 1 Overall Technical Framework

1.2.1 拓扑数据检查与修正

在供水管网的拓扑结构检查中,传统的GIS工具[如桌面GIS软件(QGIS)等]提供了一些内置功能,例如查找重复、孤立管网要素的功能,但在进行孤立要素修复时没有基于管网属性进行综合分析的方法,仍需要操作人员进行手动连接。然而,针对数据量较大的供水管网系统,QGIS及类似软件的处理方式常常面临效率和扩展性的局限性。为克服这一不足,本研究提出的供水管网拓扑修正方法基于无向图模型,通过使用DFS算法对管网的连通性进行检查与修正。该方法的核心目标是自动化检测并修

复供水管网中的重复管段、节点及孤立管段、节点,从而确保网络结构的正确性和连通性。

1) 拓扑重复的检查与修正

首先,删除重复节点和管段,以减少在连通性检查中不必要的计算(图2)。使用GeoPandas库读取Shapefile格式的供水管网数据。GeoPandas提供了高效的地理空间数据处理能力,可以直接加载和操作Shapefile文件。先通过空间距离或者精确匹配其坐标,识别并删除重复的节点,将剩下不重复的节点重新编号。接着,通过NetworkX库将加载的数据转换为无向图。在供水管网的无向图中,边的顺序

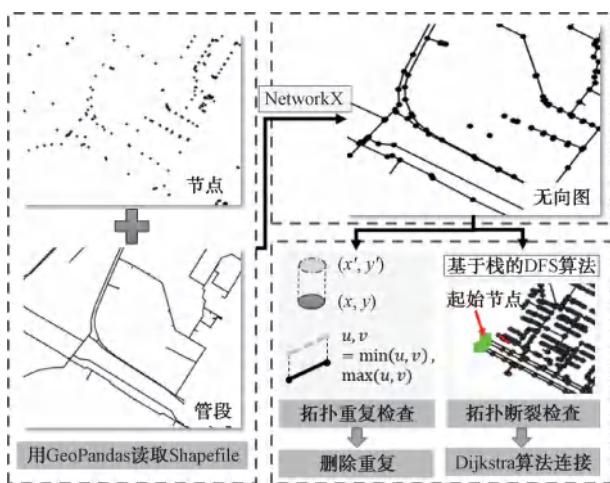


图 2 拓扑检查与修正技术流程

Fig. 2 Topology Inspection and Correction Technical Flow

对连接关系不产生影响,因此,可以删除重复的边。为此,遍历图中的所有边,将边表示为无序对,这样可以通过比较边的节点对来识别重复边。

2) 拓扑连通性的检查与修正

本文基于栈式 DFS 算法,从管网进水节点出发进行多源遍历,以实现对网络连通性的分析。首先,初始化全局访问集合 $V_{\text{visited}} = \emptyset$,依据水源点,选取进水点集合,每个进水点独立启动 DFS 遍历。进水节点要选择能够覆盖所有供水区域的进水点,可以是多个,符合多水源联合供水的拓扑特征。研究所用算法在传统 DFS 算法基础上加入栈,在栈中节点先出后进,实现边访问边记录,避免传统 DFS 算法递归深度限制。遍历进水节点后,计算孤立节点集合,如式(1),孤立节点集合等于全部节点集合减已访问的连通节点集合。若孤立节点集合非空则判定存在拓扑断裂,并提取节点关联管段集合。

$$V_{\text{isolated}} = V_{\text{total}} - V_{\text{visited}} \quad (1)$$

其中: V_{isolated} ——孤立节点集合;

V_{total} ——所有节点集合;

V_{visited} ——已访问的连通节点集合。

通过连通性分析,确定管网的连通主网络以及孤立子网络(包括单独孤立节点、2个及以上孤立节点构成的孤立网络)。再用 Dijkstra 算法计算出进水点到这些孤立子网络的最短路径并结合节点物理属性相似度进行连接,以此完成拓扑结构的修正。当所有孤立子网络完成连接,再进入连通性检查,孤立节点集合为空时结束,流程图如图 3 所示。利用

方法 Dijkstra 算法计算的最短路径时,要在连通主网络与孤立子网络之间创建虚拟连接,对每个孤立子网络筛选 n 个最短路径。再通过计算虚拟连接中涉及的连通主网络节点与孤立子网络节点属性相似度 [$\text{Sim}(v, u)$] ,进行拓扑连接。属性相似度值越大,两节点属性越相似。属性相似度计算如式(2)~式(6)。

$$\text{Sim}(v, u) = w_d \varnothing_d + w_a \varnothing_a + w_h \varnothing_h + w_m \varnothing_m \quad (2)$$

$$w_d + w_a + w_h + w_m = 1 \quad (3)$$

$$\varnothing_d = 1 - |d_v - d_u| / 500 \quad (4)$$

$$\varnothing_a = 1 - |a_v - a_u| / 50 \quad (5)$$

$$\varnothing_h = 1 - |h_v - h_u| / 2 \quad (6)$$

其中: w_d, w_a, w_h, w_m ——管径、管龄、埋深、管材的相似度权重;

\varnothing_d ——管径差异:mm;

\varnothing_a ——管龄差异:a;

\varnothing_h ——埋深差异:m;

\varnothing_m ——管材差异,相同为 1,否则为 0;

d_v ——连通主网络节点对应管径,mm;

d_u ——孤立子网络节点对应管径,mm;

a_v ——连通主网络节点对应管龄,a;

a_u ——孤立子网络节点对应管龄,a;

h_v ——连通主网络节点对应埋深,m;

h_u ——孤立子网络节点对应埋深,m。

3) 人工复检

对于检查出的拓扑问题进行人工检查,根据供水管网设计手册等资料进行核实,尤其是算法进行的自动拓扑连接的真实性。

1.2.2 属性数据检查与修正

供水管网属性数据中的缺失值是常见问题,尤其是对于管段中的管径、管龄、管材、接口形式等属性数据,缺失值的出现可能会导致管网分析结果的不准确。此外,还可能存在异常数据,一般源自测量误差、输入错误等。本文主要针对以上问题,提出了基于 Python 的自动化算法来进行属性缺失填补与异常数据修正,从而确保供水管网数据的准确性和完整性,提升数据分析的质量和可靠性。总体按以下步骤完成属性数据的自动检查与修正。

1) 异常与缺失数据的检查

供水管网的属性数据通常保存在 GIS 数据文件

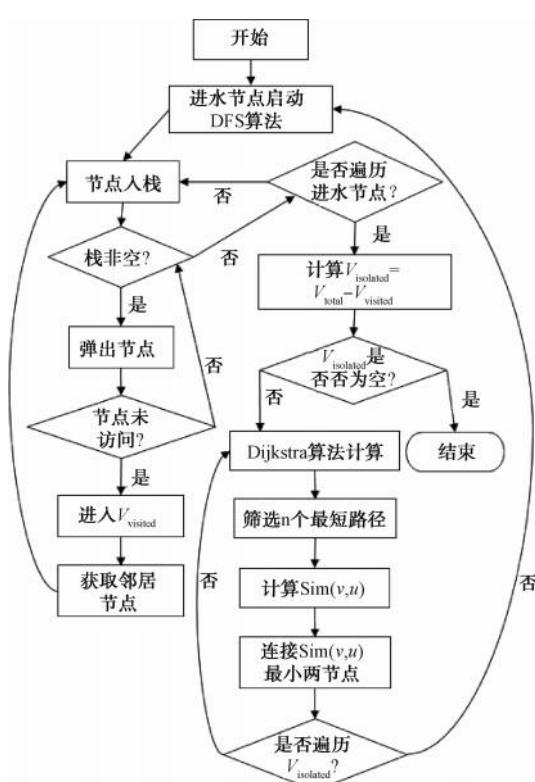


图3 拓扑检查与修正算法流程

Fig. 3 Topology Inspection and Correction Algorithm Flow

中,常见的格式包括Shapefile文件及其附带的属性数据文件(.dbf格式)。本文使用Python的dbfread模块读取.dbf格式数据,精确提取字段和记录,便于数据处理和分析。这种方法能够灵活适应不同输入格式,确保数据处理的广泛适用性。

首先,设置属性合理阈值,筛除正常范围以外的数据,主要针对0值或设计范围外管径、埋深等数据的整体筛查,即完成对全局异常值的监测。同时,筛查出空值和“NULL”这样的缺失数据,将这类异常数据与缺失数据标记,不作为下一步局部异常数据识别的考虑范围。先进行这一步的目的是防止超出事实范围和缺失的数据影响后期的自动填补工作。

再基于加权图,完成局部异常属性的检查。使用NetworkX构建加权图,权重代表管段的属性,每条边都有一个与属性相关的权重值,这样的加权图便能反映管网中各管段的物理属性和相互连接的关系,再通过建立异常判断规则,检查出异常属性数据。设定图中的某一节点为v,其相邻节点为v₁,v₂,...,v_n,相邻节点之间的管段分别为e₁,e₂,...,e_n。管段的权重为属性值w₁,w₂,...,w_n。则相邻管段之

间的差异可表示为式(7)。

$$\Delta w_{ij} = |w_i - w_j| \quad (7)$$

其中: Δw_{ij} ——相邻管段之间的权重差;

w_i和w_j——节点v_i和v_j之间连接管段的权重。

设置一个合理的阈值,用于判断管段与其相邻管段之间的属性差异是否过大,通常基于经验或统计分析确定。定义异常程度指标,来量化管段异常程度,如式(8)。当A_i的值大于1时,说明管段的属性差异显著,可能是异常的。

$$A_i = \frac{\Delta w_{ij}}{T} \quad (8)$$

其中:A_i——权重差异常指标;

T——权重差合理值。

2) 属性数据填补

在加权图的基础上,通过KNN算法和插值算法,完成对上一步检查出的异常与缺失属性数据的填补。对于分类属性(如管径、管材、管龄、接口形式等),使用KNN算法找到最近的K个邻居节点(管段),可以设定为较小的数值(如3、5),以确保修正只基于相邻的管段。并根据这些邻居的属性值进行投票,即选择最频繁的属性值作为修正值。对于埋深、高程等连续数值属性,采用IDW法来填补。IDW方法通过计算已知数据点与缺失点之间的距离加权平均,在空间上较近的已知数据点对缺失点的影响更大,从而使得填补的结果更加合理。IDW如式(9)。

$$V(X) = \frac{\sum_{i=1}^n \frac{1}{(d_i)^p} \cdot V_i}{\sum_{i=1}^n \frac{1}{(d_i)^p}} \quad (9)$$

其中:V(X)——缺失值点的估计值;

V_i——已知数据点的值;

d_i——与已知数据点的距离是插值点与第i个样本点之间的欧氏距离;

n——用于估算插值点值的样点数;

p——幂指数,一般在空间分析和等值线绘图软件中,p值取2。

当样本点分布不均匀时,可以引用调和反距离权重(anisotropic inverse distance weighting, AIDW)

方法来进行插值^[20]。

3) 人工复检

修正完数据后,需要进行人工复检。将识别的缺失与异常数据输出,参考设计资料或其他历史数据,对检查和修正的结果进行检查或抽查,对算法结果进行复核,确保数据的准确性。

属性数据检查与修正技术流程如图 4 所示。

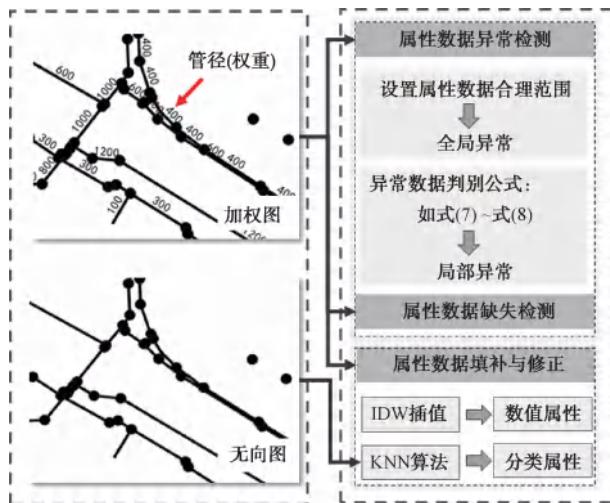


图 4 属性数据检查与修正技术流程

Fig. 4 Attribute Data Inspection and Correction Technical Flow

1.2.3 GIS 数据质量评估指标

1) 拓扑独特指数

定义一个拓扑独特指数用来评估网络中拓扑非重复的程度。拓扑独特指数越高,表示网络中的非冗余结构越多,拓扑质量越高。该指数计算如式(10)。

$$T_{\text{TUI}} = \left(1 - \frac{r_j}{t_j}\right) \times 100\% \text{ 或 } \left(1 - \frac{r_p}{t_p}\right) \times 100\% \quad (10)$$

其中: T_{TUI} —— 拓扑独特指数;

r_j 或 r_p —— 重复的节点或管段数量;

t_j 或 t_p —— 节点或管段总数量。

2) 拓扑连通指数

拓扑连通指数是衡量图中节点之间连接性的指标,通过 DFS 算法输出未连通的节点或管段数与管段总数的比值来表示,如式(11)。

$$T_{\text{TCl}} = \left(1 - \frac{u_j}{t_j}\right) \times 100\% \text{ 或 } \left(1 - \frac{u_p}{t_p}\right) \times 100\% \quad (11)$$

其中: T_{TCl} —— 拓扑连通指数;

u_j 或 u_p —— 未连通的节点或管段数量;

t_j 或 t_p —— 节点或管段总数量。

3) 数据逻辑指数

数据逻辑指数包括 GIS 逻辑属性数据总量和 GIS 逻辑属性数据异常量。GIS 属性数据总量由所评估的 GIS 数据经统计获得, GIS 逻辑属性数据的异常量经异常问题检测后统计获得。数据逻辑指数的计算如式(12)。

$$D_{\text{DLI}} = \left(1 - \frac{l_{\text{wrong}}}{l_{\text{sum}}}\right) \times 100\% \quad (12)$$

其中: D_{DLI} —— 数据逻辑指数;

l_{sum} —— GIS 逻辑属性数据总量;

l_{wrong} —— GIS 逻辑属性数据异常量。

4) 数据完整指数

数据完整指数包括 GIS 属性数据总量和 GIS 属性数据缺失量。GIS 属性数据总量和缺失量由所评估的 GIS 数据经统计获得, 属性数据总量为属性项与要素数的乘积数, 缺失量为空白属性项的个数。数据完整指数的计算如式(13)。

$$D_{\text{DWI}} = \left(1 - \frac{i_{\text{miss}}}{i_{\text{sum}}}\right) \times 100\% \quad (13)$$

其中: D_{DWI} —— 数据完整指数;

i_{miss} —— GIS 属性数据缺失量;

i_{sum} —— GIS 属性数据总量。

2 供水管网 GIS 数据质量提升技术案例分析

2.1 实例区域供水管网数据情况

为了验证供水管网 GIS 数据质量提升技术的有效性,本文选取某市的分区 Z 供水管网作为研究对象,进行相关数据的处理和分析。数据源主要由供水公司和相关工程单位提供的建设数据,以及后期通过 GIS 技术采集和更新的数据,主要包括节点、管段等 GIS 数据。节点中的属性包括节点埋深、管顶高程,管段中的属性包括管径、管材、接口形式及管段完工日期。案例的供水管网覆盖面积约为 2.34 km²,位于城市中心城区,管段、节点分别为 4 697、3 203 个,整体管网如图 5 所示。该管网包括主干道和居民区不同管网情况,具有一定代表性,涵盖了日常研究中管段与

节点层面的基本数据质量问题,具体表现:重复数据及连通性拓扑异常;管径、管材、管龄、接口形式、管顶高程以及埋深数据存在异常值与缺失现象。



图 5 某市供水管网

Fig. 5 Water Supply Pipelines Network of a City

2.2 处理结果分析

2.2.1 拓扑数据检查与修正结果

通过上述方法对研究区进行拓扑检查与修正,检查出重复管线 568 条,重复节点 502 个。如图 6 所示,这些管段重复的原因可能是数据采集时间不同或者管网改造后,数据更新不及时。

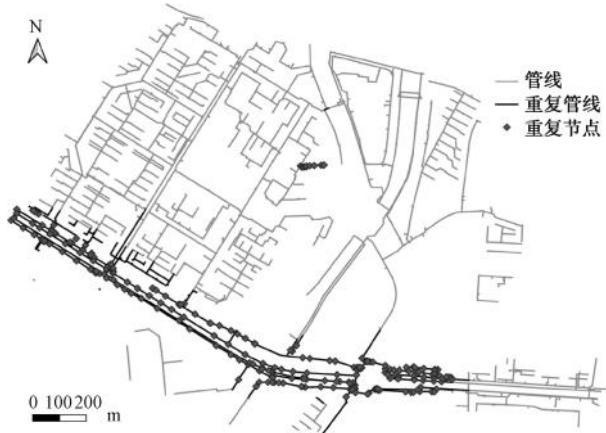


图 6 重复节点和管段检查结果

Fig. 6 Inspection Results of Duplicate Nodes and Segments

在删除重复节点和管段后,进行了管网拓扑结构的连通性自动检查与修正。结果显示,算法成功连接了 17 处拓扑断裂,涉及孤立管段共 150 根,孤立节点 158 个。经设计资料验证,确认了 15 处有效修正,成功连接节点 147 个,管段 141 条,拓扑修正错误修正率达 93.51% (表 1),这些修正扩展了有效

管网的范围。部分展示如图 7 所示(符合实际的连接标记为“T”,不符合实际的连接标记为“F”),其中有两处连接存在问题,后续可以通过实地考察进一步核实。

表 1 算法与人工方法对拓扑断裂修正准确率对比

Tab. 1 Comparison of Accuracy Rates for Algorithmic and Manual Methods in Topological Fracture Correction

指标	孤立数	算法正确修正数	人工正确修正数
节点	158	147	116
管段	150	141	105
准确率	—	93.51%	71.75%

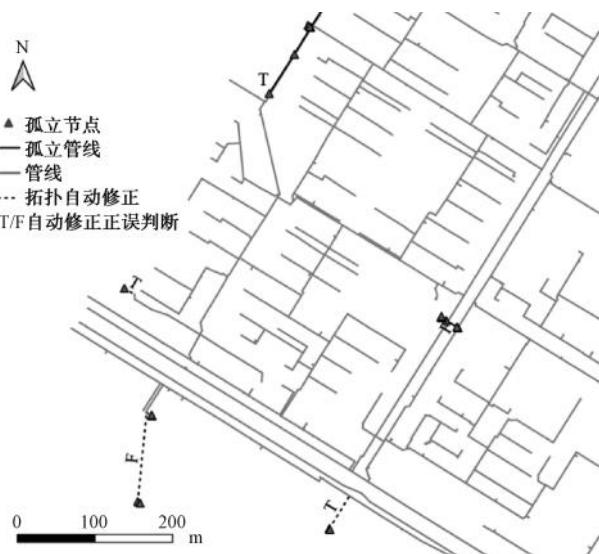


图 7 管网拓扑自动修正判断结果(部分区域)

Fig. 7 Automatic Correction Judgment Results of Pipelines Network Topology (Partial Area)

此外,研究通过对比传统 GIS 软件与所提算法的拓扑修正效果,验证了方法的优越性。试验采用 GIS 工具对案例管网进行拓扑检测,结果显示,其虽能有效识别拓扑重复及孤立子网络,但修正过程仍需人工介入。操作人员需将孤立子网络与最近的主网络节点进行物理连接,并参照设计资料进行二次验证。案例中拓扑修正算法与人工方法的管网拓扑修正情况如表 1 所示,拓扑修正算法的修正准确率达 93.51%,高出人工修正准确率 21.76%。人工修正方法准确率较低可能是因为其修正主要依赖空间距离判断,未能综合考虑管径、管材等物理属性参数,导致修正存在偏差。

2.2.2 属性数据检查与修正结果

在研究区,管网的属性数据包括埋深与管顶高

程等属性数据则存储在节点中,属于数值属性;管径、管材、接口形式、管龄等信息,这些数据存储于管段中,属于分类属性。本节将展示研究提出的算法对这些属性数据进行自动化异常值检测与缺失值填补的处理结果,同时与传统 GIS 软件处理方法进行准确性对比。首先,根据该地区的设计、施工资料对节点的埋深和管顶高程数据进行检查和填补。图 8

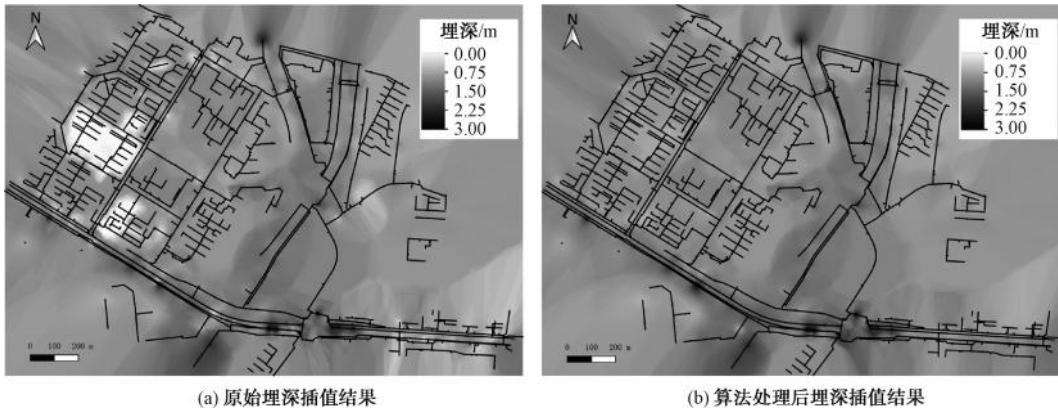


图 8 埋深处理前后对比

Fig. 8 Comparison of Processing before and after Depth Correction

本研究进一步对管径、管材、接口形式及管龄等属性参数进行算法效能验证,通过与传统 GIS 人工处理模式对属性数据填补的准确率对比,验证其优越性。如表 2 所示,算法不仅能检测出每个属性包含的缺失值,还能识别设计范围外的属性异常值和突变值,共定位 198 项属性异常与缺失问题。通过融合设计规范、施工资料等多源信息对算法修正结果进行检查,得到算法成功修正 164 项数据问题,整体修正准确率达 82.83%,较人工方法提升 17.59%。此外,通过 GIS 软件对属性数据检查出的异常值数量小于用算法检查出的异常值数量,这可能是因为 GIS 软件内置功能对局部异常识别精度不够。

表 2 算法与人工方法对异常与缺失值修正准确率对比
Tab. 2 Comparison of the Accuracy Rates of Algorithms and Manual Methods for Anomalies and Missing Values Correction

指标	本研究算法		人工方法	
	检查	修正	检查	修正
异常值	37	28	26	19
缺失值	161	136	161	103
准确率	82.83%		65.24%	

2.2.3 GIS 数据质量评估结果

根据 1.2.3 小节方法,通过拓扑独特指数、拓扑

以埋深数据为例,展示了处理前后的对比结果。数据处理后,原本数据异常与缺失导致的空白区域已被填充,不同区域的埋深变化过渡自然。结合管网施工、运维资料进行核实,数据均处于合理分布范围内,反映了处理后数据质量的显著改善。在管网 GIS 数据库中可对填补的数据进行标记,后续结合实地勘测进一步核查。

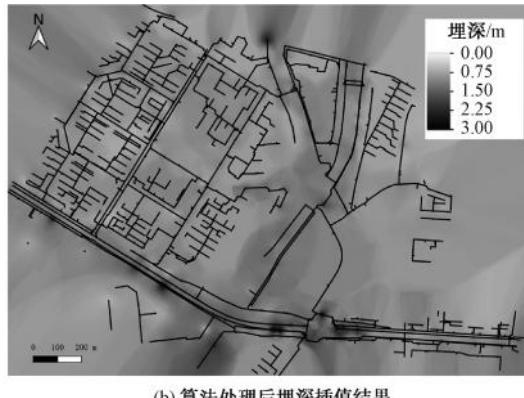


图 9 原始与修改后数据质量评估指数对比

Fig. 9 Comparison of Original and Modified Data for Quality Assessment Indices

本研究提出的供水管网 GIS 数据检查与修正方法,在提升数据质量的同时,显著优化了数据处理效率与准确率。传统 GIS 软件内置功能在拓扑数据处理中,仅能识别重复与断裂问题,修正需依赖人工基于距离逐处连接;属性数据处理中,虽然可通过字段计算器实现全局异常值与缺失值检测,但难以识别更精细的局部属性异常,后续修正仍需人工逐行处理。在包含数千管段与节点的复杂管网系统中,单纯依赖 GIS 软件工具完成数据检查与修正的效率极低。本研究设计的自动化算法,实现了拓扑与属性数据问题的全维度检测、自动修正及精准定位,大幅减少人工核查工作量,提升数据处理效能。对比试验表明,该算法在拓扑连接与属性校验的准确率上,均显著优于人工修正方法。

3 结论

本研究提出的供水管网 GIS 数据质量提升框架,通过拓扑-属性协同修正机制与自动化评估体系,系统性解决了管网数据拓扑重复、断裂及属性异常、缺失等核心问题。案例分析表明,框架中基于 DFS-Dijkstra 的拓扑检查与修正算法能够高效识别并修复拓扑错误;KNN 与 IDW 插值算法有效解决了属性数据的逻辑冲突与缺失问题,实现了属性数据的逻辑合规性优化与高效填补。通过拓扑独特指数、拓扑连通指数、数据逻辑指数和数据完整指数的量化评估,修正后的数据质量评估指数平均值由处理前的 92.36% 提升至 99.92%,表明算法显著提升了供水管网数据的整体质量。同时,相较于 GIS 软件处理方式,该框架减少了人工检查的工作量,提高了数据处理效率;拓扑修正准确率提升 21.76%,属性修正准确率提升 17.59%。该成果构建了“检查-修正-评估”的数据提升框架,为供水管网模型构建、风险评估、爆管预警及智慧水务平台建设提供了高可靠性数据基底,显著提升了供水管网 GIS 数据的质量,为城市供水管网系统的精细化运维管理奠定了坚实基础。

参考文献

- [1] 田一梅, 赵新华, 黎荣. GIS 技术在供水系统中的应用与发展[J]. 中国给水排水, 2000, 16(9): 21-23.
TIAN Y M, ZHAO X H, LI R. Application and development of GIS technology in water supply systems [J]. China Water & Wastewater, 2000, 16(9): 21-23.
- [2] 杨斌, 顾秀梅, 蒋晓君, 等. GIS 在城市供水管网信息系统中的应用[J]. 供水技术, 2007, 1(5): 47-50.
YANG B, GU X M, JIANG X J, et al. Application of GIS in the information system of urban water distribution network [J]. Water Technology, 2007, 1(5): 47-50.
- [3] 谢善斌, 袁杰, 侯金霞. 智慧水务信息化系统建设与实践[J]. 给水排水, 2018, 44(4): 134-140.
XIE S B, YUAN J, HOU J X. Construction and practice of intelligent water information systems [J]. Water & Wastewater Engineering, 2018, 44(4): 134-140.
- [4] 王博彦, 戴雄奇, 林峰, 等. 基于大数据的城市供水管网风险评估模型研究[J]. 给水排水, 2020, 46(7): 154-157.
WANG B Y, DAI X Q, LIN F, et al. Research of risk assessment model of urban water supply network base on big data [J]. Water & Wastewater Engineering, 2020, 46 (7): 154-157.
- [5] FAN X D, WANG X W, ZHANG X J, et al. Machine learning based water pipe failure prediction: The effects of engineering, geology, climate and socio-economic factors [J]. Reliability Engineering & System Safety, 2022, 219: 108185. DOI: 10.1016/j.ress.2021.108185.
- [6] ALZAROONI E, ALI T, ATABAY S, et al. GIS-based identification of locations in water distribution networks vulnerable to leakage[J]. Applied Sciences, 2023, 13(8): 4692.
- [7] 张忠贵, 王西光. 供水管网 GIS 系统与水力模型集成模式研究[C]. 上海: 供水管网建模技术论坛论文集, 2010.
ZHANG Z G, WANG X G. Integration of GIS and hydraulic models in water supply networks [C]. Shanghai: Proceedings of Water Supply Network Modeling Forum, 2010.
- [8] OLIVEIRA D P, SOIBELMAN L, GARRETT J. GIS applications for spatial analysis of water distribution pipeline breakage and condition assessment data [C]. West Lafayette: Construction Research Congress, 2012.
- [9] 杜梦航. 基于 BP 神经网络的供水管网爆管定位研究 [D]. 天津: 天津大学, 2020.
DU M H. Research on location of burst in water supply network based on BP neural network [D]. Tianjin: Tianjin University, 2020.
- [10] 李越, 杨虎, 袁卫, 等. 基于 GIS 的低功耗供水管网监测系统设计[J]. 电子设计工程, 2022, 30(18): 82-85.
LI Y, YANG H, YUAN W, et al. Design of low power water supply network monitoring system based on GIS [J]. Electronic Design Engineering, 2022, 30(18): 82-85.
- [11] 何芳, 梁艳, 谭丽红, 等. 供水管网 GIS 数据质量评价方法初探[J]. 供水技术, 2020, 14(1): 42-44.
HE F, LIANG Y, TAN L H, et al. Preliminary research on evaluation method of GIS data quality of water supply network [J]. Water Technology, 2020, 14(1): 42-44.

(下转第 194 页)

- relational learning [C] // BURSTEIN J, DORAN C, SQLORIO T. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019.
- [12] NGUYEN D Q, NGUYEN T D, NGUYEN D Q, et al. A novel embedding model for knowledge base completion based on convolutional neural network [C] // WAIKER M, JI H, STENT A. Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2018.
- [13] SCHLICHTKRULL M, KIPF T N, BLOEM P, et al. Modeling relational data with graph convolutional networks [C] // BERTINO E, GAO W, STEFFEN B, et al. Lecture notes in computer Science, 2018.
- [14] NATHANI D, CHAUHAN J, SHARMA C, et al. Learning attention-based embeddings for relation prediction in knowledge graphs [C]. Florence: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019.
- [15] XU H, BAO J, LIU W. Double-branch multi-attention based graph neural network for knowledge graph completion [C] // ROGER A, BOYD-GRABER J, OKAZAKI N. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics, 2023; 15257–15271.
- [16] BROWN T, MANN B, RYDER N, et al. Language models are few-shot learners [C] // LAROCHEHIE H, RANZATO M, HADSEII R. Advances in Neural Information Processing Systems, 2020, 33: 1877–1901.
- [17] PAN S, LUO L, WANG Y, et al. Unifying large language models and knowledge graphs: A roadmap [C]. IEEE. IEEE Transactions on Knowledge and Data Engineering, 2024.
- [18] ZHU Y Q, WANG X H, CHEN J, et al. LLMs for knowledge graph construction and reasoning: Recent capabilities and future opportunities [J]. Computer Science, 2023. DOI: 10.1007/s11280-024-01297-w.
- [19] LIU X, JI K X, FU Y C, et al. P-Tuning v2: Prompt tuning can be comparable to fine-tuning universally across scales and tasks [C] // MURESAN S, NAKOV P, VILLAVI CENCIOA. Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, 2022.
- [20] 李子豪, 冯林, 徐凌泉, 等. 基于少样本学习的知识图谱补全综述 [J]. 小型微型计算机系统, 2024, 45(12): 2838–2857.
- LI Z H, FENG L, XU L X, et al. Few-shot knowledge graph completion overview [J]. Journal of Chinese Computer Systems, 2024, 45(12): 2838–2857.

(上接第 173 页)

- [12] WU W H, PAN X Y, KANG Y K, et al. To feel the spatial: Graph neural network-based method for leakage risk assessment in water distribution networks [J]. Water, 2024, 16(14): 2017.
- [13] 王云亮, 乔庆, 韩冰. 数据驱动的供水管网安全评估模型应用实践 [J]. 给水排水, 2023, 49(7): 140–147.
- WANG Y L, QIAO Q, HAN B. Application practice of data-driven safety assessment model of water supply pipeline network [J]. Water & Wastewater Engineering, 2023, 49(7): 140–147.
- [14] 李杉杉. 基于机器学习的市政管网运维风险评估 [D]. 哈尔滨: 哈尔滨工业大学, 2021.
- LI S S. Risk assessment of municipal pipenetwork operation and maintenance based on machine learning [D]. Harbin: Harbin Institute of Technology, 2021.
- [15] 应路瑶. 基于水力模型的供水管网独立计量分区优化方法研究 [D]. 哈尔滨: 哈尔滨工业大学, 2022.
- YING L Y. Optimization study on district metered area by hydraulic model inwater distribution system [D]. Harbin: Harbin Institute of Technology, 2022.
- [16] 朱鹏飞. 基于神经网络的供水管网爆管风险评估模型研究 [D]. 上海: 上海财经大学, 2023.

- ZHU P F. Risk assessment model for pipe bursts in water supply networks using neural networks [D]. Shanghai: Shanghai University of Finance and Economics, 2023.
- [17] 李振博. 基于改进差分进化的供水管网 DMA 分区研究 [D]. 昆明: 昆明理工大学, 2023.
- LI Z B. Research on district metered areas of water distribution network based on improved differential evolution [D]. Kunming: Kunming University of Science and Technology, 2023.
- [18] WANG M, MA Y, WU C, et al. A data-driven online model system for urban water pipeline network based on GIS [C]. Wuhan: International Conference on Electronics Information and Emergency Communication, 2024.
- [19] ABDULSAMAD A A, ABDULRAZZAQ K A. Calibration and analysis of the potable water network in the Al-Yarmouk region employing WaterGEMS and GIS [J]. Journal of the Mechanical Behavior of Materials, 2022, 31(1): 298–305.
- [20] 李正泉, 吴尧祥. 顾及方向遮蔽性的反距离权重插值法 [J]. 测绘学报, 2015, 44(1): 91–98.
- LI Z Q, WU Y X. Inverse distance weighted interpolation involving position shading [J]. Acta Geodaetica et Cartographica Sinica, 2015, 44(1): 91–98.