



基于村庄分散度的农村生活污水治理模式量化分析

李桂兰¹, 严滢^{1,2}, 李鹏宇^{2,3}, 王紫轩^{1,2}, 周子玉^{1,2}, 郑天龙^{2,3,✉}, 刘建国^{1,✉}, 刘俊新^{2,3}

1. 内蒙古工业大学资源与环境工程学院, 环境污染控制与修复内蒙古自治区高等学校重点实验室, 呼和浩特 010051; 2. 中国科学院生态环境研究中心水污染控制实验室, 北京 100085; 3. 中国科学院大学, 北京 100049

摘 要 随着农村地区生活水平的提升, 生活污水所含污染物成分趋于复杂化, 未经处理就地排放对生态环境影响不容小觑。目前, 农村生活污水治理模式种类多样, 但决策方法缺乏科学依据。以内蒙古地区 82 个村落为研究对象。以自然村落为最小研究单元, 利用村落宅基地占比、村落长宽比、周长、极限距离、坡度 0°~8°占比和坡度标准差等指标, 构建了居住密度决策树 (准确率 76%) 以及地势决策树 (准确率 88%)。决策树所得结果通过二分类概率单位回归 (Probit) 分析 (准确率 87%), 确定其村落集中或分散情况, 进而选择适宜的村落污水治理模式 (集中模式、纳管模式、分散模式、拉运模式、就地利用模式)。农村生活污水治理模式决策方法的确立决策指标的量化, 为促进农村生活污水治理体系的发展、提升污水治理的有效性具有积极意义。

关键词 村庄分散度; 农村生活污水; 机器学习; 决策树; 治理模式

农村生活污水治理是改善农村人居环境, 提升美丽乡村建设水平, 实现乡村振兴战略的重要途径。依据最新调查统计数据, 目前中国农村地区的生活污水产量每年约为 1.3×10^{10} t^[1]。大部分农村地区没有污水处理设施, 大量生活污水直接排入地表水体或河流, 直接影响饮水安全, 导致土壤肥力下降, 粮食产量降低^[2]。国家相继出台了《农村人居环境整治三年行动方案》^[3]《农村人居环境整治提升五年行动方案 (2021—2025 年)》^[4]《农业农村污染治理攻坚战行动方案 (2021—2025 年)》^[5]《“十四五”土壤、地下水和农村生态环境保护规划》^[6]。截至 2022 年, 全国农村生活污水治理率仅为 31%。随着国家对农村污水处理的高度重视, 确定适合的治理模式来解决农村生活污水问题显得尤为重要^[7]。

农村生活污水处理受到多种因素的影响, 主要包括人口因素、自然因素和经济因素^[8]。其中, 人口因素包括人口数量、人口居住分散或聚集程度等; 自然因素主要为地形地势、地区环境敏感程度、水系特征等; 经济因素主要包括建设投资、运行维护成本等。农村人口基数差异较大, 部分村落中青年村民常年在外务工, 人口流失造成住宅长期空置^[9]。人口数量直接影响村庄污水排放量以及处理规模。依据农村居民人口数量以及居住聚落可分为大型聚落, 中型聚落, 小型聚落, 独院聚落, 复杂的居住聚落形态可能会增加其收集处理难度^[10]。内蒙古地区特有“六牧四农”的发展格局, 形成诸多农牧交错带, 其聚落空间既具有普遍性, 又因其特殊的交错地理位置多位于山地而具有独特性^[11]。内蒙古作为一个农牧业大省, 其农牧区占全区土地面积的 60% 以上。农牧区居民大多被农田环绕或依托牧区草场分散居住, 其形成聚落复杂多样。目前生活污水处理多为集中模式, 通过管网系统将区域内收集的废水统一输送到集中的设备进行处理。农村生活污水产生量少, 同时居住布局的密度影响其管道输送距离, 收集距离较长易造成收集不完全^[12]、财政支出较大等问题。因此, 有必要依据村庄居住密度探索适合农村的生活污水治理模式。此外, 在不同自然、经济、社会等环境因素的综合作用下, 农村居民点空间分布也会表现出显著的地域分布特征^[13]。已有研究表明, 坡度 0°~8°的区域农村居民居住比重最高, 8°~25°次之, 25°以上极少^[14-17]。地形坡度直接影响管道铺设的距离及

收稿日期: 2023-11-07; 录用日期: 2023-11-14

基金项目: 国家自然科学基金资助项目 (51838013), 内蒙古自治区“草原英才”工程创新创业人才团队

第一作者: 李桂兰 (1974—), 女, 硕士, 副教授, 465646162@qq.com; ✉通信作者: 郑天龙 (1988—), 男, 博士, 副研究员, tlzheng@rcees.ac.cn; 刘建国 (1980—), 男, 博士, 教授, jgliu_yc@163.com

难易程度。因此,除居住密度外坡度对生活污水治理模式的影响不可忽略。

目前,有关农村生活污水治理模式的研究呈现多样性和多维性。汪泽锋等^[18]依据村庄附近是否布设城镇污水收集管网、村庄居民分散程度和地形,将农村生活污水治理模式分为纳管处理、集中处理和分散处理三种模式。揭维慧等^[19]结合研究区域的调研情况,从已建污水处理设施规模以及建设运行方式角度,将农村污水处理设施划分为 3 种模式,即乡镇级污水处理设施、中心村庄建立的小型污水处理设施和分散式农户改厕。而高生旺等^[20]等通过对 22 个省份 396 个典型农村生活污水治理情况调研总结出规模与排放导向相结合的治理模式,主要有集中处理-达标排放、集中处理-资源化利用、分散处理-达标排放、分散处理-就地利用 4 大类。易嘉静等^[21]在此基础上,基于当地案例进行分析,丰富了生活污水治理模式将其分为市政纳管治理模式、村级集中治理模式、村级生态治理模式、分散处理就地利用模式和灰水存储清运处理模式。目前相关研究已经得到广泛关注,形成了一个相对丰富的研究领域。但其结论仅止步于对现状的总结归纳,无法从指标量化的角度针对拟治理村落的治理模式进行较科学的决策。

本研究选取位于内蒙古自治区的 82 个村落作为研究对象,将决策农村污水治理模式的指标体系量化并使用决策树进行居住密度和地形分类模型的构建,可视化地分析了农村生活污水治理模式的选择依据,提高了农村污水治理决策的科学性和严谨性。弥补了我国在农村生活污水治理模式方面研究的不足,为农村人居环境综合整治工作提供理论依据和技术支撑。

1 材料与方法

本研究基于自然村尺度的居住密度、地势和其他相关因素,构建居住密度分析方法,确定适宜此地区的生活污水治理模式。决策框架如图 1 所示,通过对不同条件下治理模式的比较和评估,可以有效地制定出科学合理的污水治理策略,以提高农村生活污水治理效果和水环境保护水平。

1.1 研究对象

内蒙古自治区作为我国北部边疆,农畜产品资源多样性丰富,人均耕地面积 3 590 m²,是全国人均耕地面积的 3 倍。有“塞外米粮川”之称是国家农业开发的重点地区,其畜牧业综合生产能力位于全国五大牧区之首。本区地貌以高原型为主,大部分地区海拔超过 1 000 m。除了高原地形,该地区还包括山地、丘陵和平原等其他地貌类型。农村生活污水未经治理随意排放易对农田等造成潜在影响。内蒙古农村地区的生活习惯,污水水量冬季排放量低,夏季排放量高;且日变化波动差异较大,早中晚时间段排放量高,其余时间段几乎无污水排放。目前,内蒙古地区污水治理率仅为 16.85%^[22]远低于全国平均农村污水治理率。未处理的农村污水直接排入环境当中直接威胁农村居民健康。

在内蒙古农牧地区以及平原、山地等地区选取了 82 个村落作为研究样本。首先,使用 Bigemap GIS Office 30.0.0.0 软件对村落底图进行提取,以获取村落的空间信息。进一步,借助 ARCGIS 10.7 软件对村落的院落数量(个)、宅基地面积(m²)、村落长宽(m)、周长(m)等要素进行提取。这些要素的提取工作经过 ARCGIS 软件进行处理,以确保数据的准确性和一致性。随后,利用 Excel 软件对提取的要素进行运算,得到居住密度相关指标的具体数值。

同时,为了获得内蒙古地区的地形数据,本研究采集了地理数据空间云的 Digital Elevation Model (DEM) 数据。DEM 数据用于描述地形的高程变化情况,能够提供有关地势和地貌的重要信息。使用 ARCGIS 10.7 软件对 DEM 数据进行分析后,可以获得与地形相关的数据,如地势坡度、高程差等。

综合以上步骤,本研究通过对 82 个村落的底图提取和分析,以及对地理数据空间云的 DEM 数据的应用,得到了具体的村落居住密度相关指标和地形相关数据。这些指标和数据将在后续的研究中用于居住密度分析和生活污水治理模式的确定。依托现有研究将其治理模式分为 5 种。纳管模式:铺设管道直接接入市政

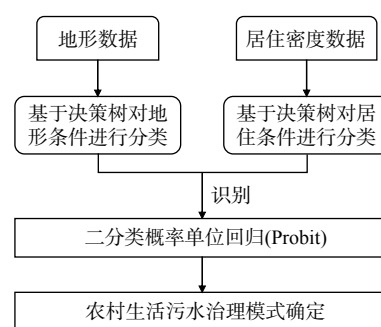


图 1 村庄分散度分类及识别模型

Fig. 1 Classification and recognition model of village dispersion

污水管网进行统一集中处理。集中处理模式：村内统一建设污水处理站进行处理。拉运模式：人口较少距离城镇管网较近的，宜采取户收集拉运模式，定期将污水运至城镇污水处理厂集中处理。分散处理模式：人口居住相对分散或受地势限制，环境有一定要求的区域宜选择分散式处理模式。就地利用模式：适用于人口居住分散度高，地广人稀，具备较好资源化利用条件的地区，生活污水优先考虑就地、就近、就农资源化^[23]。经决策模型构建将确定治理模式的指标量化，科学的确定农村污水治理模式。

1.2 分析方法

在常见的分类算法中，包括决策树(分类树)、随机森林、支持向量机和神经网络分类等。这些模型在实践中被广泛使用，其中决策树算法因其能够直观地展示详细分类过程而备受关注。相比之下，其他分类模型更偏向于黑箱模型，即难以解释其预测过程，但能够取得更好的预测性能。决策树算法由 LeoBrei-man 等学者在 1984 年提出^[24]，并发展出 C4.5、C5.0 等算法模型。决策树的表现形式为二叉树，模型构建包括生长和剪枝 2 个过程^[25]。

对于一个给定的数据集，具体计算方法如下：假设 $p(i|t)$ 表示给定节点 t 中属于类 i 的记录所占的比例，在二分类问题中，任意结点的类分布都可以记作 $(p0, p1)$ ，其中 $p1=1-p0$ 那么对最优的划分度量常见的有 3 种。

信息熵(香农熵)见式(1)。

$$\text{Entropy}(t) = - \sum_{i=0}^{c-1} p(i|t) \log_2 p(i|t) \quad (1)$$

基尼系数见式(2)。

$$\text{Gini}(t) = 1 - \sum_{i=0}^{c-1} [p(i|t)]^2 \quad (2)$$

分类误差见式(3)。

$$\text{Classificationerror}(t) = 1 - \max_i [p(i|t)] \quad (3)$$

决策树方法的优点包括易解释性、计算效率高、能够处理高维数据和非线性关系等。同时该方法能够通过图形表达，更直观地对分类过程给出依据，相较于其他黑箱决策树模型有更好地可视性和可解释性^[26]。

1.3 指标选取

本研究从自然村居住密度及地势 2 方面进行量化决策。根据研究目的，我们将居住密度类型分为 5 类，地势类型分为 3 类。

1) 宅基地占比。村庄宅基地占比如式(4)所示是指在一个村庄中宅基地所占的比例。宅基地是指村庄居民建造院落的面积(如图2)。

$$A = \frac{S_{\text{宅基地}}}{S_{\text{村庄}}} \quad (4)$$

2) 村落长宽比。村落的长宽比(C)是指村庄的长度与宽度之间的比例关系。它是用来描述村庄形状的一个指标。长度通常是指村庄从一端到另一端的距离，可以被视为村庄在一个方向上的延伸。而宽度则是指村庄在垂直于长度方向的宽度。具体计算公式如式(5)所示。

$$C = \frac{L_{\text{宽}}}{L_{\text{长}}} \quad (5)$$

3) 周长。村落周长是指村庄或村落的周围边界的总长度。用于一个描述村落大小和形状的指标。潜在涵盖村庄的规模大小以及村庄轮廓形状的复杂程度，用于反应村庄布局结构。



图2 指标参数取值

Fig. 2 Index Parameter Values

4) 极限距离。通过村落的水量以及管道管径,分析得出村落管道铺设的极限距离,极限距离范围内的居民纳入村庄一并治理。以极限距离为界,范围以外居民不纳入村落区域范围内。具体计算公式如式(6)所示。

$$d = 9.393 \frac{\left(\frac{Q}{S}\right)^{0.6244}}{D^{1.2253}} \quad (6)$$

式中: d 为临界距离, km; Q 为水量, $\text{m}^3 \cdot \text{day}^{-1}$; S 为村庄面积, km^2 ; D 为管径, mm。

5) 坡度标准差。基于坡度的地表粗糙度指数,使用单元格斜率完成计算,并根据局部表面坡度变化进而评估地形的平缓度或复杂度。该算法通过计算移动窗口的平均斜率与其内部单个单元格斜率之间的差值来消除趋势引起的误差^[27]。具体计算公式^[28]如式(7)所示。

$$\delta s = \sqrt{\frac{1}{N^2} \sum_{i=1}^{N^2} (S_i - \bar{S})^2} \quad (7)$$

式中: δs 为坡度标准差; S_i 为移动窗口内第 i 个像元的坡度, °; $\bar{S}$ 为移动窗口内像元的平均坡度, °; N 为移动窗口的宽度(单元格数)。

6) 坡度 $0^\circ \sim 8^\circ$ 占比情况。对于坡度的划分,根据《土地利用现状调查技术规程》和国际地理学会的标准,我们可以将其划分为 $0^\circ \sim 2^\circ$ (极缓坡), $2^\circ \sim 5^\circ$ (缓坡), $5^\circ \sim 8^\circ$ (中等坡), $8^\circ \sim 15^\circ$ (斜坡), $15^\circ \sim 25^\circ$ (陡坡), $25^\circ \sim 35^\circ$ (急坡), $35^\circ \sim 90^\circ$ (险坡)。因此,本研究通过分析村落坡度 $0^\circ \sim 8^\circ$ 占比确定村落地势情况。

1.4 特征分类

本研究基于村落居民居住密度情况,对其院落布局进行归纳整理将其分为五类,进而确定其适合的污水处理模式。其中星型村落较为特殊,多出现在牧区,相邻两户之间相隔数百米,整体村落宅基地占比 $< 5\%$,建设集中型污水处理设施收集难度较大,收集的污水水量无法满足集中型污水处理设施的正常运行,因此,单户处理更为适宜。此外,考虑到地形对星型村落的影响较小。本研究将宅基地占比小于 5% 的村落归类为星型村落,有助于优化决策树分析,更好地对村庄的院落布局进行分类和总结。具体村庄内部分布如图3所示。

1) 密集型。村庄中的居民住房和院落布局相对紧密,表现出一种相对集中的空间结构,具有一个明确的



图3 村落形态示意图

Fig. 3 Schematic diagram of village morphology

中心点。较适宜进行集中治理，在全村范围内铺设管道后，纳入中央污水处理设施进行处理。

2) 块状集中。村庄中的居民住房和院落布局发展出了多个相对独立的聚集区域，呈现多个明显的中心点。适宜在各个独立聚集区域范围内分别进行集中治理。

3) 密集与零散相结合型。大多数村民的居住分布相对集中，形成了一个明确的中心区域，同时也存在少部分居民选择在相对零散的区域居住，相对分散于中心区域之外。较适宜于中心区域建立污水处理设施，将集中居住的居民纳入管网，距离较远的零散户可进行分散治理。

4) 分散型。村落中的居民分布较为零散，院落之间的距离相隔几十米甚至上百米。村庄中无明显的中心聚集点。较适宜单户或临近几户共同建立小型分散式污水处理设施，进行分散处理。

5) 星型。村落中的居民散居，院落相互之间距离较远。村落没有明显的中心聚集点，每个居民的住宅都相对独立。适宜在单户范围内进行分散治理，处理后可直接就地利用资源化。

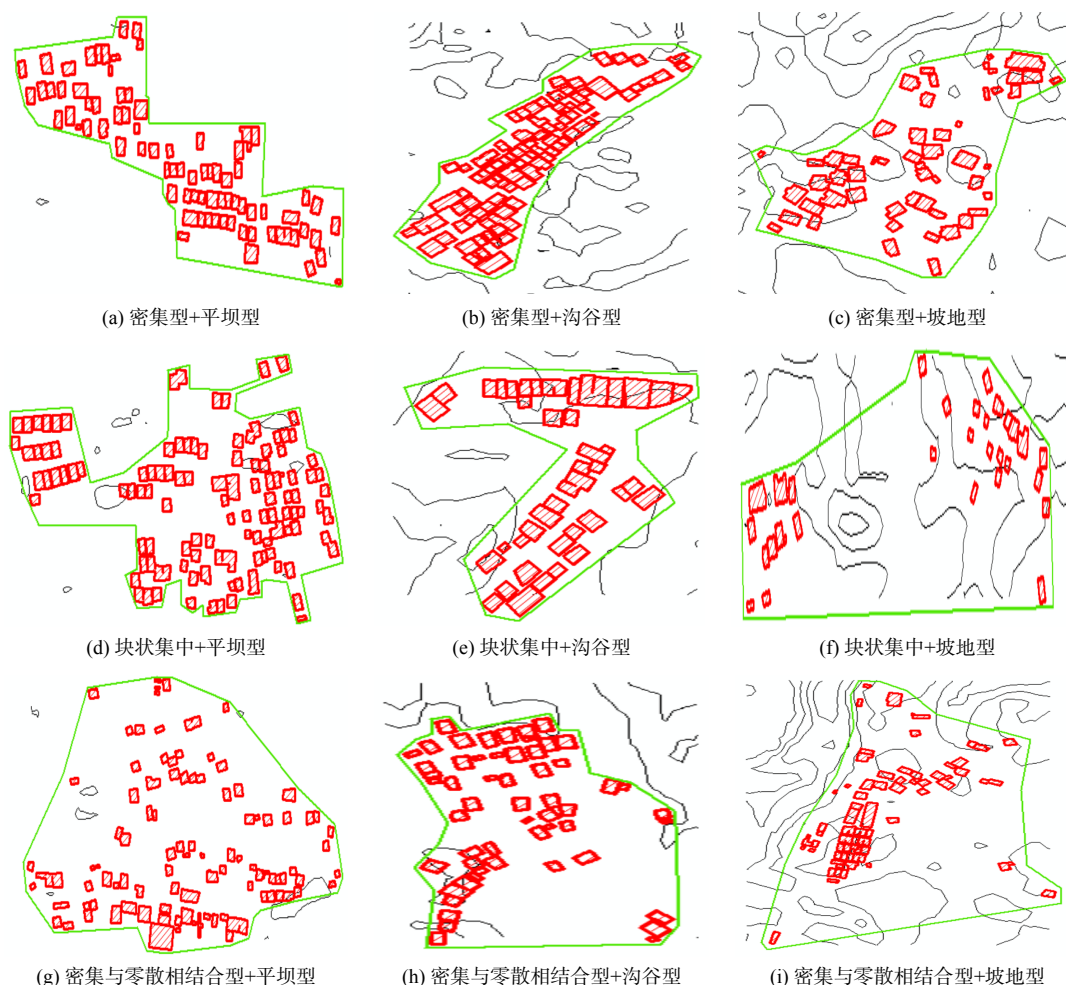
基于居住密度确定其属性的基础上，可初步确定其治理模式。但村落所在地地形起伏较大使管道铺设的实际距离增加、污水收集率降低。地形坡度过大增加了管道铺设的难度以及建设费用^[29]。管道排水坡度的增加导致管道寿命降低和污水倒流进而影响管道系统的正常运行。平坦地区铺设管道难度低，污水可以通过管道就近进行集中处理。而在山区或地势崎岖的地区，管道铺设难度较大，难以进行统一收集需要采取分散排放或使用地理式处理设施等适应地形的方式。将其分为以下3类。

6) 平坝型。地势起伏较小，较为平坦，有利于管道铺设。

7) 沟谷型。地势起伏沟壑不平，村庄地形构成复杂，导致处理设施的布置和管道网络的设计困难。

8) 坡地型。地势从高处逐渐向低处倾斜，具有较大的高差和坡度，地势倾斜导致管道布置不便，使得排水困难。

村庄经居住密度和地形分析后可归为12种类型，如图4所示。



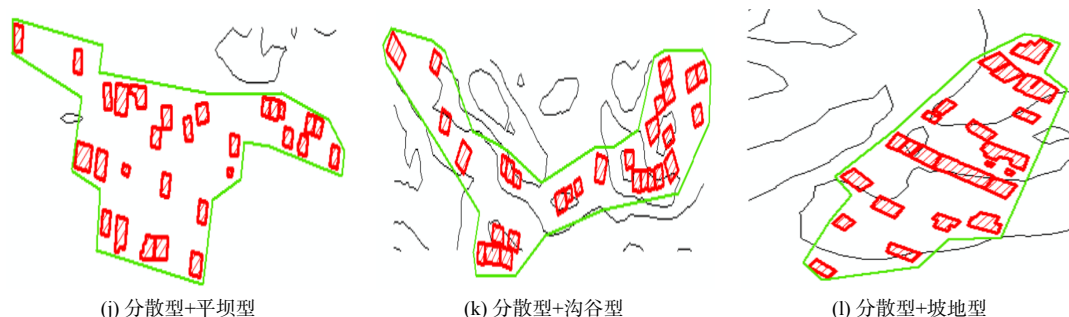


图4 村庄居住密度及地形分析类型

Fig. 4 Village residential density and terrain analysis type

2 结果与讨论

根据上述样本数据,用决策树对居住密度和地形分别进行建模,居住密度决策树将自变量(宅基地比例、长宽比、周长、极限距离)和因变量(密集型、块状集中、密集与零散相结合型、分散型)纳入模型,模型的准确率为76.47%,精确率(综合)为80.39%,召回率(综合)为76.47%,f1-score(综合)为0.77。模型预测结果与实际观测分析结果对比如图5所示。

地形决策树将自变量(坡度 $0^{\circ}\sim 8^{\circ}$ 占比、坡度标准差)和因变量(平坝型、坡地型、沟谷型)纳入决策树模型,准确率为88.00%,精确率(综合)为96.21%,召回率(综合)为96.00%,f1-score(综合)为0.96。模型预测结果与实际观测分析结果对比如图6所示。运用决策树构建的模型,在对村庄居住离散度判别同实际观测结果基本相符,说明了模型评价的有效性。

基于2个决策树模型结果将其治理类型进行赋值,令密集型=1;块状集中=2;密集与零散相结合型=3;分散型=4。采用二分类概率单位回归(Probit)确定农村生活污水情况为集中或分散进而确定其具体治理模式。基于模型构建方法,不同地势类型的推导公式如式(8)~式(10)所示。

$$\text{平坝型: } P = 2.417 - 0.805 \text{类型} \quad (8)$$

$$\text{沟谷型: } P = -0.218 - 0.805 \text{类型} \quad (9)$$

$$\text{坡地型: } P = 1.195 - 0.805 \text{类型} \quad (10)$$

由表1可见,模型结果较理想,准确率达到87.80%。由上式可知, $P>1$ 时为集中,其人口居住密度较高且地势平缓,建议经管道收集后进行集中治理; $P<1$ 时为分散,宜采用分散治理。

受试者工作特征曲线(receiver operating characteristic curve, ROC)曲线是一种常用于评估分类模型性能的工具。它以真正例率(true positive rate, TPR)为纵轴,以假正例率(false positive rate, FPR)为横轴进行绘制。通过ROC曲线如图7所示,分析可知该模型的ROC曲线下面积(AUC=0.899, $P<0.01$, 95%CI:-2.695-2.084)而AUC越接近于1,表示模型的性能越好。由此可得,本模型可信度较高。

其农村生活污水治理模式选择方法流程如图8所示。基于居住密度决策树以及地势决策树分析后对村落进行分类,分类结果经二分类概率单

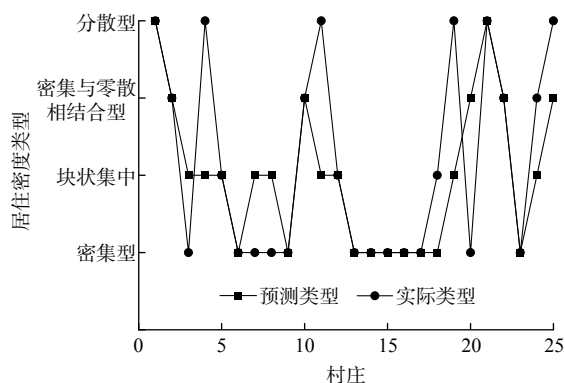


图5 居住密度类型判定对比图

Fig. 5 Comparison chart of residential density type determination

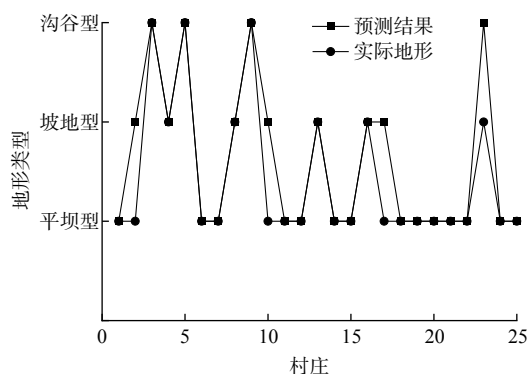


图6 地形类型判定对比图

Fig. 6 Comparison chart of terrain type determination

位回归分析后确定其村落属于集中类型村落或分散类型村落。将村落居住密度及地形信息量化后经决策树及二分类概率单位回归分析后，为农村污水治理模式提供较科学的选择依据。

为直观的分析该方法在农村分散度分类中的效果，将逻辑回归算法和支持向量机 (support vector machine, SVM) 算法分别建模进行比较。以相同的输入样本数据为基础，对 3 种建模应用中的分类结果进行比较，结果如表 2 所示。决策树分类模型表现出较高的分类准确率，在决策评价中显示出较强的分类能力和抗干扰能力，该模型具备较好的泛化性能和稳定性。相比之下，逻辑回归分类模型的分类准确率次之，而支持向量机分类模型较差。此评价结果表明，基于决策树算法的居住密度定类及地形定类适合对村庄分散度的判定，可以对农村污水治理模式做出合理有效的评价。

表 1 模型结果参数
Table 1 Model Result Parameters

准确率	召回率	精确率	F1	AUC
87.80%	87.80%	88.10%	87.80%	89.90%

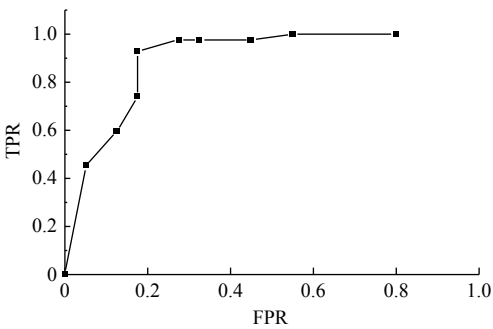


图 7 ROC 曲线

Fig. 7 Receiver operating characteristic

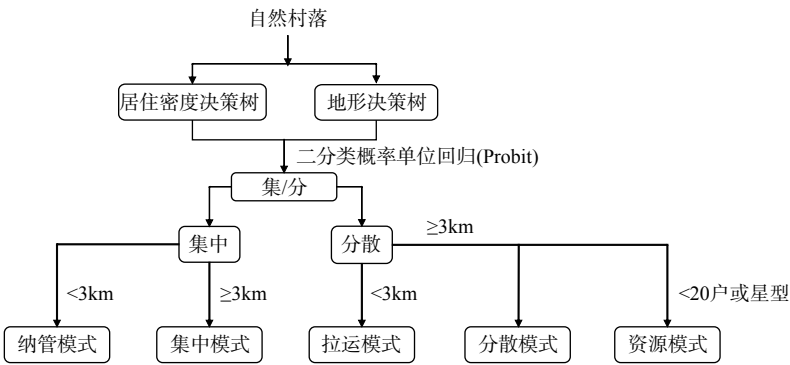


图 8 生活污水治理模型

Fig. 8 Domestic sewage treatment model

3 结论

基于内蒙古地区 82 个村庄的居住密度及地形数据，利用决策树模型在分类方法中的适用性，建立了基于村庄分散度确定生活污水治理模式的分类模型。建立了村庄居住密度决策树模型 (准确率为 76.47%) 和地势分类决策树模型 (准确率：88.00%)，并与二分类概率单位回归 (Probit) 模型 (准确率为 87.80%) 相结合确定农村生活污水治理模式。将 30.00% 的数据样本带入模型测试，得到的准确率较高，以上模型具有良好的稳定性和较高的准确性。此外，通过与逻辑回归和支持向量机模型进行对比得出，使用决策树模型得到的居住密度类别及地形类别判定结果要优于逻辑回归和支持向量机两种算法。决策树模型量化了确定村庄分散度的指标，提高了农村污水治理模式选择的效率，为农村污水治理提供了科学依据。

表 2 分类准确率比较

Table 2 Comparison of classification accuracy

分类类型	决策树	逻辑回归	支持向量机
居住密度类型	76.47%	52.00%	48.00%
地形类型	88.00%	72.00%	64.00%

参考文献

[1] GUI Z Y, WEN J H, FU L, et al. Analysis on Mode and Benefit of Resource Utilization of Rural Sewage in a Typical Chinese City[J]. Water, 2023, 15: 11.
[2] LI Z H. Distributed treatment of rural environmental wastewater by artificial ecological geographic information system[J]. Environmental Research, 2021, 112572.
[3] 中共中央办公厅, 国务院办公厅. 农村人居环境整治三年行动方案[S/OL]. [2023-08-01]. http://www.gov.cn/zhengce/2018-02/05/content_5264056.htm.

- [4] 中共中央国务院. 中共中央国务院印发《农村人居环境整治提升五年行动方案(2021-2025年)》[EB/OL]. [2023-08-01]. http://www.gov.cn/zhengce/2021-12/05/content_5655984.htm.
- [5] 生态环境部, 农业农村部, 住房和城乡建设部, 等. 关于印发《农业农村污染治理攻坚战行动方案(2021—2025年)》的通知(环土壤[2022]8号)[EB/OL]. [2023-08-01]. https://www.mee.gov.cn/xxgk/xxgk03/202201/t20220129_968575.html.
- [6] 生态环境部, 国家发展和改革委员会, 财政部, 等. 关于印发《“十四五”土壤、地下水和农村生态环境保护规划》的通知(环土壤[2021]120号)[EB/OL]. [2023-08-01]. https://www.mee.gov.cn/xxgk/xxgk03/202112/t20211231_965900.html.
- [7] BI X Q, GUO L J. Treatment process of rural domestic sewage based on small purification tank technology[C]. 2nd International Conference on Air Pollution Environmental, 2020, 450.
- [8] 付浩, 闫海, 邱长浩. 人口密集地区农村生活污水治理若干问题探讨[J]. 给水排水, 2020, 56(9): 9-14.
- [9] LI J, YU R X. A study on villages of the planning of water infrastructure planning: 25 villages in Shaanxi Province as an example[J]. Aqua-Water Infrastructure Ecosystems and Society, 2022, 72: 19-31.
- [10] 郝小雨. 陕南秦巴山区乡村聚落空间格局特征及优化研究[D]. 长安大学, 2022.
- [11] 张凯, 马明. 内蒙古农牧交错带山地聚落空间形态研究[J]. 地域研究与开发, 2019, 38(6): 165-170.
- [12] 陈秀丽. 高质量高标准建设背景下农村生活污水治理模式探讨[J]. 给水排水, 2022, 58(12): 69-75.
- [13] 李君. 农户居住空间演变及区位选择研究[D]. 河南大学, 2009.
- [14] 胡燕, 胡玉福, 邓良基, 等. 基于GIS的石角镇农村居民点空间分布特征分析[J]. 中国农学通报, 2010, 26(16): 315-320.
- [15] 李国敬. 基于GIS的渭河流域居民点空间分布特征研究[J]. 测绘与空间地理信息, 2022, 45(3): 130-133.
- [16] 陈丽. 基于GIS的县域农村居民点整理研究[D]. 南京大学, 2013.
- [17] 范红梅, 王秋兵, 边振兴. 基于GIS技术的宽甸县居民点空间分布特征分析[J]. 西南师范大学学报(自然科学版), 2008(2): 99-103.
- [18] 汪泽锋, 操家顺, 王超越, 等. 农村生活污水处理技术与模式研究情况[J]. 应用化工, 2022, 51(9): 2669-2674+2680.
- [19] 揭维慧. 金寨县农村生活污水治理模式研究[D]. 合肥工业大学, 2023.
- [20] 高生旺, 黄治平, 夏训峰, 等. 农村生活污水治理调研及对策建议[J]. 农业资源与环境学报, 2022, 39(2): 276-282.
- [21] 易嘉静. 农村生活污水治理思路及应用案例分析[J]. 节能与环保, 2022(10): 85-86.
- [22] 李铁霄, 郑天龙, 杨晓霞, 等. 西北地区农村生活污水排放标准对比及分析[J/OL]. 工业水处理: 1-18. <https://doi.org/10.19965/j.cnki.iwt.2022-1140>.
- [23] 中国城镇供水排水协会. 中国城镇水务行业年度发展报告[M]. 北京: 中国建筑工业出版社, 2022: 306-308.
- [24] SAFAVIAN S R, LANDGREBE D. A Survey of Decision Tree Classifier Methodology[J]. Ieee Transactions on Systems Man and Cybernetics. 1991.21: 660-674.
- [25] PANHALKAR A R, DOYE D D. A novel approach to build accurate and diverse decision tree forest[J]. Evolutionary intelligence, 2022, 15: 439-453.
- [26] 贾册, 陈臻, 韩梅. 基于决策树模型的区域PM_{2.5}污染管控时空识别——以关中地区为例[J]. 干旱区研究, 2022, 39(4): 1056-1065.
- [27] 王泓, 黄伟亮, 项闻, 等. 滑坡粗糙度年代学方法——以甘肃舟曲地区泄流坡和锁儿头滑坡为例[J]. 地球科学与环境学报, 2022, 44(6): 1037-1047.
- [28] FRANKEL K L, DOLAN J F. Characterizing Arid Region Alluvial Fan Surface Roughness with Airborne Laser Swath Mapping Digital Topographic Data[J]. Journal of Geophysical Research:Earth Surface, 2007, 112(F2): 000644.
- [29] 程方奎, 汪晨晨, 温仓祥, 等. 太湖流域农村生活污水产污特征及治理模式分析[J]. 东南大学学报(自然科学版), 2022, 52(3): 578-585.

(责任编辑: 金曙光)

Quantitative analysis of rural domestic sewage treatment mode based on village dispersion

LI Guilan¹, YAN Ying^{1,2}, LI Pengyu^{2,3}, WANG Zixuan^{1,2}, ZHOU Ziyu^{1,2}, ZHENG Tianlong^{2,3,*}, LIU Jianguo^{1,*}, LIU Junxin^{2,3}

1. College of Resources and Environment, Inner Mongolia University of Technology, Hohhot, Inner Mongolia Autonomous Region 010051, China; 2. Water Pollution Control Laboratory, Ecological Environment Research Center, Chinese Academy of Sciences, Beijing 100085, China; 3. University of Chinese Academy of Sciences, Beijing 100049, China

*Corresponding author, E-mail: tlzheng@rcees.ac.cn; jgliu_ycx@163.com

Abstract With the improvement of living standards in rural areas, the composition of pollutants contained in domestic sewage tends to be complicated, and the impact of untreated on-site discharge on the ecological environment cannot be underestimated. At present, there are various types of rural domestic sewage treatment modes, but the decision-making methods lack scientific basis. This paper took 82 villages in Inner Mongolia as the research object. Taking the natural village as the smallest research unit, the residential density decision tree (accuracy : 76%) and the terrain decision tree (accuracy : 88%) were constructed by using the proportion of village homestead, the ratio of village length to width, perimeter, limit distance, the proportion of slope 0°~8 ° and the standard deviation of slope. The results of the decision tree were analyzed by the binary classification probability unit regression (Probit) analysis (accuracy rate : 87%) to determine the concentration or dispersion of the villages, and then select the appropriate village sewage treatment mode (centralized mode, receiving mode, decentralized mode, pulling mode, local utilization mode). The establishment of the decision-making method of rural domestic sewage treatment model has constructed and quantified the decision-making indicators, which is of positive significance for promoting the development of rural domestic sewage treatment system and improving the effectiveness of sewage treatment.

Keywords village dispersion; rural domestic sewage; machine learning; decision tree; governance model