基于大语言模型的环保举报数据提升管理效率研究

王希铭[1,2]，钟丽锦[3]，丁隆真[1,2]，吕广丰[3]，齐兴育[3]，胡清[1,2*]

（1. 南方科技大学土壤污染防治与安全全国重点实验室，广东深圳 518055；2. 南方科技大学环境科学与工程学院，广东深圳 518055；3. 北京环丁环保大数据研究院，北京 100083）

【摘 要】环保举报是公众参与环境治理的关键渠道之一，公众通过环保举报提供了大量非结构化的文本数据。然而，传统结构化标签分析方法难以有效利用文本信息，限制了从环保举报中发现潜在环境问题的能力。本研究基于2016—2021年全国567 985条环保举报数据，引入DeepSeek大语言模型对非结构文本进行挖掘，系统比较其与结构化标签对污染类型识别的表现。研究发现：①非结构文本可提升污染类型识别准确性，非结构文本性能较高（F1得分中位数：0.92），优于结构化标签（F1得分中位数：0.50）；②非结构文本可降低易感知污染类型的感知偏差（关联规则置信度差异下降9.6%），更真实地反映公众关切的实际环境问题；③非结构文本对时空特征更敏感，有利于发现公众关切变化的早期信号。案例分析表明，该方法对养殖废水和工业废气的识别时间分别提前约4年与2年，并揭示出公众对噪声类问题日益上升的关注趋势。本研究提供了一种快速识别环保举报非结构化文本中公众隐性关切环境问题的基于大语言模型的新方法，可辅助管理部门快速响应公众需求、推动精细化管理。

【关键词】环保举报；大语言模型；文本挖掘；公众参与；公众科学；环境管理

【中图分类号】X32；TP391　　【文章编号】1674-6252（2025）06-0106-10

【文献标识码】A　　【DOI】10.16868/j.cnki.1674-6252.2025.06.106

引言

《“十四五”生态环境监测规划》提出研究构建群众接受度高、反映获得感强的生态环境质量综合评价方法，使评价结果与人民群众感受更加一致[1]。随着居民环境意识提升，公众环境关切数量呈明显上升趋势[2]，公众参与已成为现代环境治理体系的重要组成部分[3]。据统计，2019年我国环保举报数量比2013年增长989.6%[4]。这些数据被用于中央生态环保督察、“接诉即办”等监管机制，已展现出较好的应用价值[5,6]。然而，公众参与在环境治理体系中仍处于结构性边缘地位[7]，如何更好将其融入决策前端仍是研究者面临的挑战[8]。

环境管理部门曾在20世纪90年代提出将环保举报替代直接监测的设想，并使用环保举报中结构化的污染类型标签（以下简称“结构化标签”）进行可行性研究，结果表明其对危害性较大却不易察觉的污染物缺乏敏感性[9]。近年来，学界基于公众感知（Public Perception）探索了环保举报在识别污染源[10]、评估环境质量[11]、完善环境标准[12]及优化监管资源配置[13]等场景的应用表现，结果显示

作者简介：王希铭（1989—），男，博士研究生，研究方向为大数据与生态环境，E-mail：xim_wang@126.com。

*责任作者：胡清（1964—），女，博士，教授，博士生导师，研究方向为大数据与生态环境、城市环境工程、污染场地绿色可持续修复与管理、生态环境发展战略与政策标准研究、环境健康、绿色环保及循环技术等，E-mail：huq@sustech.edu.cn。

结构化标签存在一定局限性，如污染类型的分类准确性不足、感知偏差、样本代表性偏差等问题。目前，环保举报作为辅助环境监测的数据可靠性仍受到怀疑[14]。

现有环保举报分析方法依赖结构化标签，而对描述环境事件的非结构化文本（简称“非结构文本”）缺乏有效分析方法。公众科学（Citizen Science）研究指出，非结构文本可弥补结构化标签在信息捕捉上的局限[15]，人工智能的发展为挖掘非结构文本信息提供了潜在技术路径[16]。将大语言模型（Large Language Model）与环保举报结合，有望提升后者的数据质量与价值，进而推动生态环境监测的方法论创新。本研究旨在探索一种能够解析海量且具有描述性的公众关切，并将其转化为监管可用数据的方法，从而唤醒处于“沉睡”状态的公众非结构文本数据资源，为提升环境治理现代化水平提供参考。

1 实验设计

公众感知将人类社会视为一个分布式传感器网络，基于个体行为数据挖掘公众群体诉求[17]，并通过这种方式预警公共事件[18]。环保举报通过结构化标签实现公众感知，但该机制存在数据偏差：①标签识别偏差：当举报人遇到标签术语与个体认知不匹配的情况时，可能误选标签[19]；②感知偏差：公众主要依赖感官性状（如嗅觉、视觉、听觉等）识别污染[20]，但该过程易受个体认知、社会网络、文化背景、媒体等因素影响[21]，可能导致公众感知与科学测量结果出现差异[22]；③代表性偏差：标签系统早期建设不完善导致样本缺失、地区覆盖不足，影响数据代表性[23]。在环保举报过程中，举报人需提供非结构化的环境事件描述文本（如举报对象、污染问题等）。非结构文本中的语义状态信息，可与结构化标签构成互补与互证关系。然而，目前尚缺乏对环保举报中非结构文本与结构化标签进行系统性比较的研究。

基于以上分析开展实验设计，如图1所示：①提取环保举报中的结构化标签和每条标签样本所对应的非结构文本；②利用大语言模型识别非结构文本所反映的污染类型标签（简称“非结构文本标签”）；③通过人工标注的基准标签，评估结构化标签与非结构文本的污染类型识别准确性；④利用关联规则分析，评估二者的感知偏差；⑤通过案例分析，比较三种污染类型标签数量在时空特征敏感性的差异。

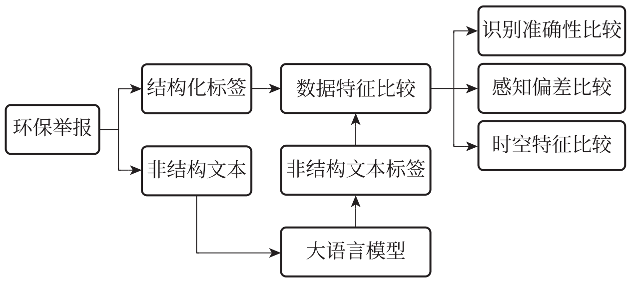

图1 研究路线

2 材料与方法

2.1 数据来源和样本选取

本研究基于生态环境部“全国生态环境投诉举报平台”，获取了2016年1月1日—2021年12月31日的环保举报样本567 985条。该数据主要变量包括：时间戳、经纬度、污染类型、环境事件描述。污染类型为定类多选，是结构化标签的数据来源，环境事件描述是非结构文本的数据来源。

2.2 分析方法

2.2.1 基于污染类型标签计算样本量百分比

由于举报人在一个举报件中可选择多个污染类型标签，研究通过式（1）计算污染类型的样本量百分比，以衡量其样本数量特征：

$$P_i = \left(\frac{N_i}{N_t}\right) \times 100\% \tag{1}$$

式中，P_i表示第i种污染类型的样本量百分比；N_i表示第i种污染类型标签出现的频数；N_t表示举报总件数。研究发现污染类型的样本数量N_i符合幂律分布特征，即少数头部类别占据了绝大多数样本，导致平均值易受极端值影响。因此，为更稳健地描述N_i分布特征，研究选用中位数表征样本量及其百分比的中心趋势，以四分位距反映其离散程度。基于此，在后续分析中，采用样本量百分比的中位数反映数量特征随时间变化趋势，采用样本量百分比的极差反映其在区域分布上的异

质性。

2.2.2 基于大语言模型识别非结构文本标签

文本挖掘已广泛运用于公众环境关切文本分析[24,25]，然而传统方法存在语义解析不足、小样本分类表现偏弱、依赖人工标注等局限性[26,27]，较难解析口语化表达中的隐含信息，在多污染、未知污染类型识别中表现不佳。目前，大语言模型已较好地解决了上述不足[28]，但尚缺乏将其专门应用于环保举报非结构文本的研究。我国学者提出的 DeepSeek-V3 开源模型与当前主流大语言模型（Claude、Gemini、GPT 等）相比，分类准确性较高、使用成本较低[29]，可兼顾效率与数据安全。综上，研究通过本地部署的 DeepSeek-V3 构建文本主题分类模型，生成非结构文本标签，评估了每种污染类型的精确率、召回率、F1 分数（F1-score）。以各类别 F1 分数的中位数代表平均分类水平，以宏平均精确率—召回率曲线下面积评估模型在应对污染类型不均衡时的综合表现。

2.2.3 基于非结构文本抽样构建基准标签

为评估识别准确性，本研究构建了一套污染类型基准标签体系。其构建基于监督式机器学习专家标注流程[30]：其一，从非结构文本中对所有污染类型进行等比例随机抽样；其二，邀请环境科学领域的专业标注员对样本进行标注与数据脱敏；其三，对存在标注冲突的样本进行校准与剔除。经此流程，共获得 4012 条污染类型基准标签（每种污染类型标注 100 条以上）。基准标签用于评估结构化标签与非结构文本的污染类型识别准确性。

2.2.4 污染类型的关联规则分析

为探究非结构文本对特定污染类型感知偏差的调节作用，本研究基于关联规则（Association Rule），构建了关联规则置信度指标[31]。关联规则 $A \rightarrow B$ 的置信度定义为当污染类型 A 出现时，另一种污染类型 B 在同一举报件中出现的条件概率，如式（2）所示。该指标量化了举报人描述污染类型 A 时，同时提及污染类型 B 的可能性：

$$\operatorname{conf}(A \rightarrow B)=P(B \mid A) \tag{2}$$

当规则 $A \rightarrow B$ 与反向规则 $B \rightarrow A$ 的置信度不相等时，这种非对称性反映出举报人表达污染类型时的偏好[31]，即感知偏差。为量化此偏差，研究通过式（3）计算两条规则置信度之差，得到关联规则置信度差异指标：

$$\Delta_{A-B}=\operatorname{conf}(A \rightarrow B)-\operatorname{conf}(B \rightarrow A) \tag{3}$$

当 $\Delta_{A-B}>0$ 时，表明举报人对 B（相较于 A）存在正向偏差（反之则为负向偏差），$|\Delta_{A-B}|$ 反映偏差强度。

设所有污染类型的集合为 $\boldsymbol{S}$，特定污染类型为其子集 $\boldsymbol{X}$，其余污染类型为 $\boldsymbol{Y}=\boldsymbol{S} \setminus \boldsymbol{X}$，对于所有满足 $B \in \boldsymbol{X}$ 且 $A \in \boldsymbol{Y}$ 的污染类型配对 (A,B)，研究分别计算其在非结构文本中的置信度差异 $\Delta_{A-B}^{(\text{unstr})}$ 与结构化标签中的置信度差异 $\Delta_{A-B}^{(\text{str})}$，构建以下两个置信度差异绝对值的集合，如式（4）所示：

$$\begin{aligned}\mathcal{D}_{\text{unstr}}&=\{|\Delta_{A-B}^{(\text{unstr})}| \mid \forall A \in \boldsymbol{Y}, B \in \boldsymbol{X}\}\\ \mathcal{D}_{\text{str}}&=\{|\Delta_{A-B}^{(\text{str})}| \mid \forall A \in \boldsymbol{Y}, B \in \boldsymbol{X}\}\end{aligned} \tag{4}$$

集合 $\mathcal{D}$ 中的每个元素，表征了特定污染类型 B 与任一其他污染类型 A 配对所呈现的偏差强度。通过比较 $\mathcal{D}_{\text{unstr}}$ 与 $\mathcal{D}_{\text{str}}$ 分布的集中趋势，可以评估非结构文本与结构化标签对特定污染类型偏差强度的差异。若 $\mathcal{D}_{\text{unstr}}$ 的中位数经配对样本 Wilcoxon 符号秩检验证实显著小于 $\mathcal{D}_{\text{str}}$ 的中位数（$p<0.05$），则表明在非结构文本中，公众对特定污染类型的感知偏差强度较结构化标签显著减弱[32]。

3 结果与分析

3.1 环保举报数据的规律和特点

研究分析了环保举报中污染类型的规律和特点：①样本分布不均衡。图 2a 展示了结构化标签中，样本量占比大于 0.1% 的 30 种污染类型的分布及其累计百分比。图例所示 10 种污染类型占样本总量的 87.6%，其中恶臭异味和烟粉尘占总样本量的 40.8%，反映出污染类型数量上的集中分布特点。从空间分布看，广东、河南、山东、江苏四个省份的样本量合计占比 42.6%，显示出地域不均衡特点。为进一步揭示样本空间聚集模式，图 2b 采用统一流形逼近与投影（UMAP）随机抽取了 30 万条结构化标签样本，基于其经纬度和污染类型进行降维，发现样本主要集中于城市并以城市群为聚

集特征。其中，恶臭异味与烟粉尘（红色点）集中于城镇化发展水平较高地区，二者在粤港澳大湾区（20.8%、13.8%）、长三角城市群（14.1%，12.0%）、京津冀城市群（7.8%，10.0%）的样本量百分比位居前列。②非结构文本标签数量更多，较结构化标签增加 25.8%，效应量分析表明 7 种污染类型（工业噪声、建设施工噪声、工业废气、生活废水、社会生活娱乐噪声、黑臭水体、养殖废水）的数量变化值得关注（Cohen's h 绝对值 > 0.1），如图 2c 所示。③非结构文本标签中污染类型标签共现数量更多。结构化标签中，同一举报件中涉及 2 种及以上污染类型占总样本量的 29.9%，非结构文本标签增加至 58.8%。以上结果表明，非结构文本能够提供更丰富的数据特征。同时，分析发现从城市群的尺度来表达污染类型的区域分布特征，可能较行政区划更具代表性。

3.2 非结构文本提升污染类型识别准确性

基于基准标签的评估结果表明，非结构文本对污染类型的识别性能优于结构化标签。如图 3a 所示，在与 30 种污染类型的基准标签对比中，非结构文本的平均识别准确率（F1 分数中位数为 0.92，四分位距为 0.86～0.97）明显高于结构化标签（F1 分数中位数为 0.50，四分位距：0.28～0.71），说明非结构文本能更准确地提取举报人描述的环境事件信息。图 3（b）显示，非结构文本识别能力已接近人工标注水平（宏平均精确率—召回率曲线下面积为 0.90），而结构化标签的识别表现较弱（宏平均精确率—召回率曲线下面积为 0.56），说明在污染类型样本量不均衡场景下，基于大语言模型的识别方法展现出较好的稳定性，初步显示出替代部分人工识别的潜力。

3.3 非结构文本降低污染类型感知偏差

恶臭异味、烟粉尘等被视为当前公众关切最强烈的污染类型 [33]，其具有易感知的特点，容易引起环境舆情 [34]。图 3c 将恶臭异味与烟粉尘设为特定污染类型，计算其相对其余 28 种污染类型的感知偏差（关联规则置信度中位数为 36.2%，四分位距为 21.3%～48.1%）高于反向关联规则（关联规则置信度中位数为 4.7%，四分位距为 1.5%～8.0%），存在显著差异（Wilcoxon 符号秩检验，$p < 0.001$），表明结构化标签中恶臭异味、烟粉尘存在明显的正向偏差。图 3d 将恶臭异味与烟粉尘设为特定污染

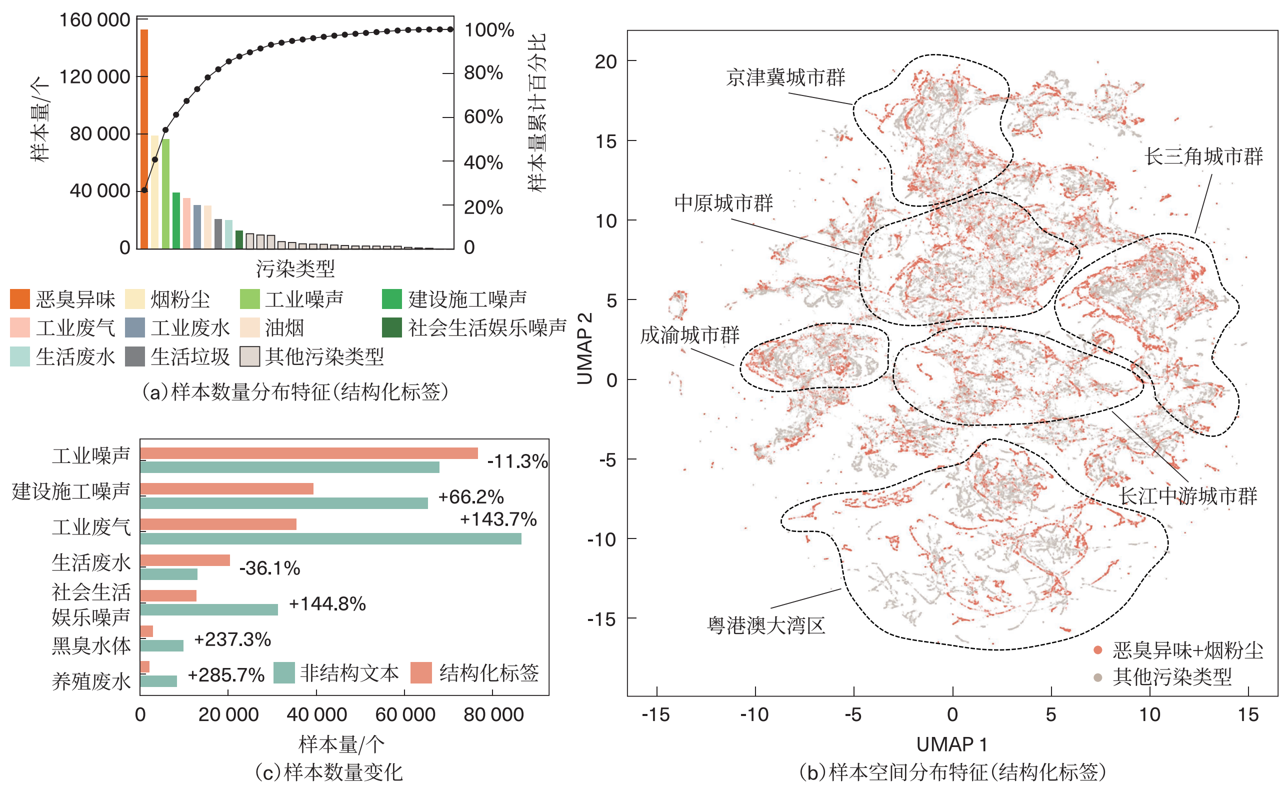

图 2　污染类型的数量和空间分布特征

类型，比较其在结构化标签与非结构文本中的感知偏差，发现后者的感知偏差显著下降（关联规则置信度差异为 −9.6%，四分位距为 −17.1%～−2.3%，Wilcoxon 符号秩检验，$p < 0.001$），证明非结构文本能够调节恶臭异味、烟粉尘的感知偏差。图 4 分别将 19 种污染类型设为特定污染类型（其余 11 种污染类型因样本量较少不满足检验要求而被排除），比较其在结构化标签与非结构文本中的感知偏差，发现恶臭异味、烟粉尘、工业噪声的正向偏差强度显著下降，工业废气的正向偏差强度显著上升，矿渣尾矿、黑臭水体、地下水污染、养殖废水、机动车尾气的负向偏差强度显著下降。

结构化标签易受举报人对污染类型的偏好影响，导致易感知污染类型（如空气污染、噪声污染）与不易感知污染类型（如固废污染、土壤污染）之间出现样本量差异，可能使紧迫性、隐蔽性污染类

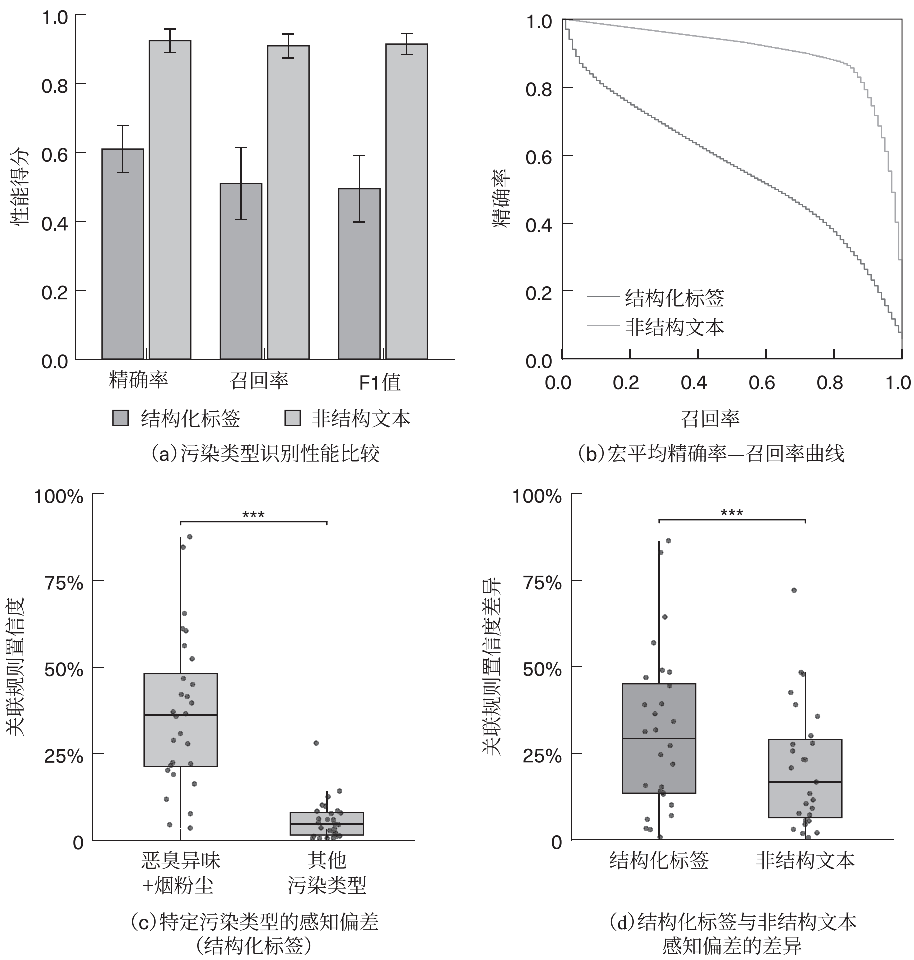

图 3　污染类型的识别准确性和感知偏差

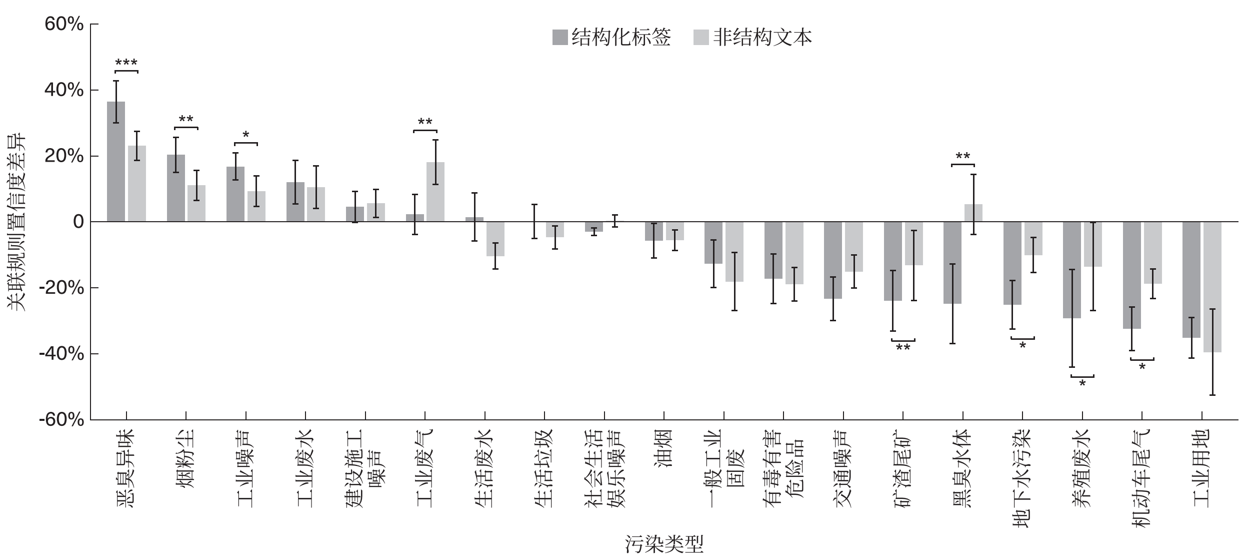

图 4　污染类型的关联规则置信度差异

注：图中使用标准误（Standard Error of the Mean）表示误差。研究采用 t 检验方法检验某种污染类型在结构化标签和非结构文本的关联规则置信度差异是否存在显著变化，* 代表存在显著变化（* 表示 $p < 0.05$，** 表示 $p < 0.01$）。

型数量被低估，对决策优先级及监管资源配置产生不成比例的影响[35]。非结构文本降低了易感知污染类型的感知偏差，更真实地反映公众环境关切。

3.4 非结构文本提高污染类型时空特征敏感性

环保举报主管部门通过回溯污染类型标签的时空特征（如逐年上升或某地区占比较大），辅助政策分析。由于结构化标签中恶臭异味、烟粉尘等污染类型存在感知偏差，致使其不易被感知，但可能存在较大危害的污染类型数量被低估，如工业废气和养殖废水等。研究发现，非结构文本数量特征在时间、空间维度的敏感性，有利于更早发现环境问题。以养殖废水、工业废气、社会生活娱乐噪声等为例进行分析。

3.4.1 养殖废水

表 1 所示结构化标签中恶臭异味相对养殖废水的感知偏差（关联规则置信度差异：86.4%）高于非结构文本（关联规则置信度差异：66.2%）。与基准标签相比，结构化标签的召回率（0.22）较低，非结构文本的召回率（0.97）较高，说明举报人在非结构文本中描述了养殖废水问题，在结构化标签中倾向使用恶臭异味标签表达该问题，仅依靠结构化标签会导致养殖废水标签大量遗漏。从时间维度看，图 5a 所示非结构文本的标签数量（样本量百分比中位数为 3.6%，四分位距为 2.4%～5.3%，2016 年）显著高于结构化标签（样本量百分比中位数为 1.3%，四分位距为 1.0%～1.7%，2020 年；Wilcoxon 秩和检验，（p<0.001），呈显著上升趋势（Mann-Kendall 检验，p < 0.05，2016—2018 年），表明养殖废水标签数量的全国平均水平早在 2016 年就已达到并超过结构化标签（2020 年），提前 4 年出现标签数量上升特

表 1　结构化标签与非结构文本在识别三种污染类型时的表现差异

污染类型	组别	识别准确性		感知偏差	时空特征敏感性	
		精确率	召回率	关联规则置信度差异	样本量百分比中位数	样本量百分比极差
养殖废水	结构化标签	0.85	0.22	86.4/%	1.3/%	2.0/%
	非结构文本	0.97	0.97	66.2/%	6.2/%	8.3/%
工业废气	结构化标签	0.76	0.25	46.9/%	13.7/%	14.0/%
	非结构文本	0.92	0.88	15.3/%	30.7/%	22.8/%
社会生活娱乐噪声	结构化标签	0.78	0.75	− 7.6/%	4.8/%	9.2/%
	非结构文本	0.87	0.95	9.0/%	7.6/%	16.0/%

注：关联规则置信度差异计算中，研究选取恶臭异味—养殖废水、恶臭异味—工业废气、油烟—社会生活娱乐噪声作为分析案例。样本量百分比极差计算中，研究基于结构化标签首次出现的年份进行比较：养殖废水（2020 年）、工业废气（2018 年）、社会生活娱乐噪声（2018 年）。

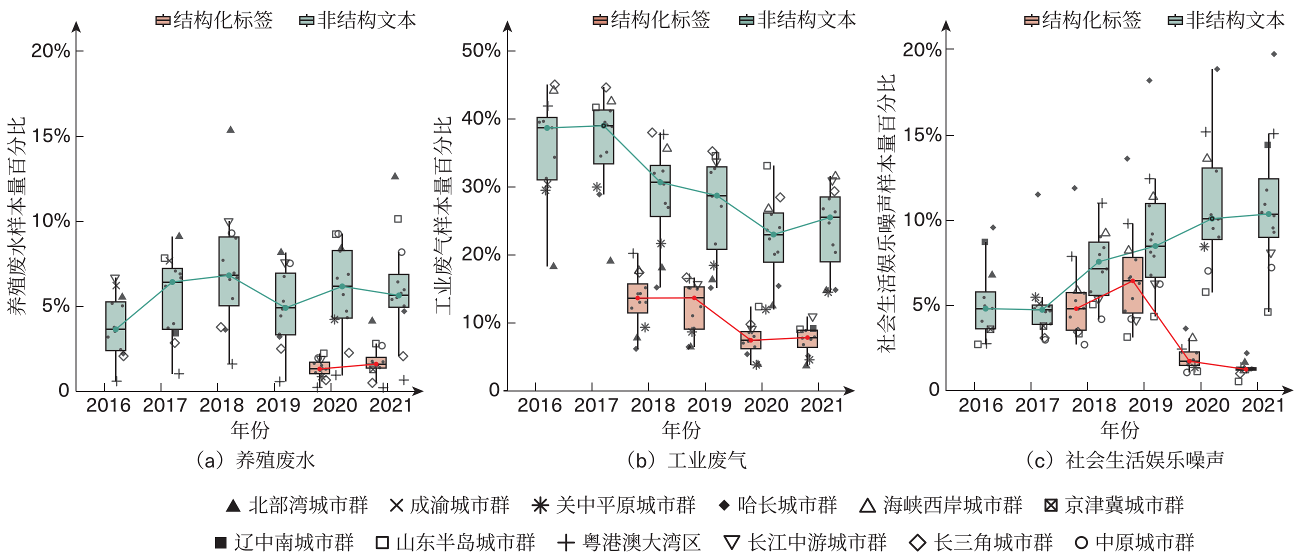

图 5　三种污染类型在我国主要城市群的样本量百分比随时间变化

征。从空间维度看，表 1 所示非结构文本在城市群之间的污染类型空间分布，与结构化标签相比差异较大（样本量百分比极差提升 6.3%）。另外，如图 5a 所示，2016—2018 年北部湾、长江中游、成渝城市群（样本量百分比中位数为 6.7%，四分位距为 6.6%～9.1%）显著高于粤港澳大湾区、长三角城市群（样本量百分比中位数为 1.8%，四分位距为 1.2%～2.6%），空间异质性突出。

3.4.2 工业废气

如表 1 所示，恶臭异味相对工业废气的偏好明显降低（关联规则置信度差异下降 31.6%），识别准确率明显提升（召回率上升 0.63），说明举报人在非结构文本中描述了工业废气问题，在结构化标签中倾向使用恶臭异味标签表达该问题，仅依靠结构化标签会导致工业废气标签大量遗漏。图 5b 所示的非结构文本标签数量（样本量百分比中位数为 38.7%，四分位距为 31.1%～40.2%，2016 年）显著高于结构化标签（样本量百分比中位数为 13.7%，四分位距为 11.5%～15.7%，2018 年），呈现显著下降趋势（Mann-Kendall 检验，$p < 0.05$，2016—2019 年），表明非结构文本提前 2 年发现工业废气标签数量下降特征。非结构文本也表现出较强的空间敏感性（样本量百分比极差提升 8.8%）。2016—2019 年，长三角、海峡西岸及山东半岛城市群的标签数量（样本量百分比中位数为 38.8%，四分位距为 35.1%～43.0%）显著高于北部湾、哈长城市群（样本量百分比中位数为 22.1%，四分位距为 17.6%～29.1%）。

3.4.3 社会生活娱乐噪声

如表 1 所示，油烟相对社会生活娱乐噪声的感知偏差较低且无明显变化（油烟与社会生活娱乐噪声标签在环保举报中存在较高关联性），识别准确性提升不明显。图 5c 所示的非结构文本标签数量（样本量百分比中位数为 4.8%，四分位距为 3.9%～6.4%，2016 年）与结构化标签（样本量百分比中位数为 4.8%，四分位距为 3.8%～6.7%，2018 年）无显著差异，表明社会生活娱乐噪声标签数量上升趋势及时被发现。从空间维度看，2018—2019 年，粤港澳大湾区、海峡西岸城市群标签数量（样本量百分比中位数为 11.2%，四分位距为 10.6%～11.7%，哈长城市群作为异常值剔除）显著高于京津冀、山东半岛及长江中游城市群（样本量百分比中位数为 5.2%，四分位距为 4.5%～6.0%）。

3.5 非结构文本更好地揭示公众需求

为比较非结构文本和结构化标签在整体表达公众关切的差异性，图 6 分析了主要污染类型样本量排名逐年变化（排名按当年样本量百分比的降序排列）。如图 6a 所示，结构化标签的排名年度波动较大，部分污染类型存在缺失。结构化标签在 2017—2020 年存在调整，其中工业废气、生活

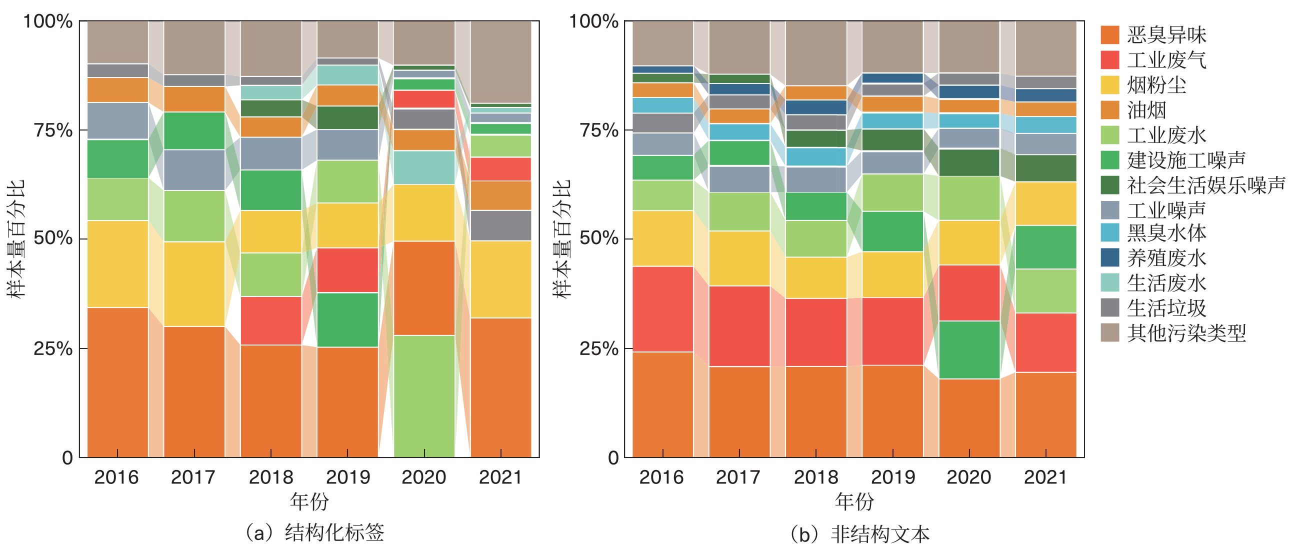

图 6　主要污染类型的全国样本量百分比排名逐年变化

废水、社会生活娱乐噪声标签创建于2018年，黑臭水体、养殖废水标签创建于2020年。环保举报中7.9%的样本存在污染类型标签缺失，其中噪声污染、水污染分别缺失19.9%、14.7%，直接导致2020—2021年生活废水、养殖废水、社会生活娱乐噪声的标签数量偏低，影响了数据的时序可比性。反观非结构文本，其完整捕捉了上述缺失的公众关切。图6b所示的非结构文本中，社会生活娱乐噪声、建设施工噪声的排名上升较快，烟粉尘、工业废水、养殖废水、黑臭水体的排名先上升后逐年下降，而工业废气、恶臭异味的排名未明显变化。非结构文本清晰地揭示出公众对噪声等污染日益上升的关注，提供了更连续、更准确的追踪手段。

传统监测网络往往无法覆盖所有地域和环境参数，公众更接近环境事件源头，环保举报以低成本、大范围的优势，提供环境问题的预警并填补常规监测的空白[36]。研究发现：①基于大语言模型的非结构文本挖掘，提升了环保举报中污染类型识别的准确性。非结构文本的污染类型识别准确率（F1得分中位数：0.92）明显优于结构化标签（F1得分中位数：0.50）。②降低了公众在表达易感知污染类型时的感知偏差，对恶臭异味与烟粉尘相对其余污染类型的关联规则置信度差异下降9.6%。③提高了污染类型的时空特征敏感性。时间维度，非结构文本较结构化标签更早识别养殖废水（提前4年）和工业废气（提前2年）的数量特征。空间维度，其识别结果在不同城市群间表现出更大异质性（对养殖废水、工业废气的样本量百分比极差分别扩大6.3%和8.8%）。

研究通过三种污染类型，具体分析了非结构文本如何更好地揭示公众需求：其一，自2016年起，养殖废水标签数量已呈现明显上升趋势，其中北部湾、长江中游城市群的空间异质性显著高于长三角、粤港澳大湾区城市群。该特征与我国2015年“水十条”实施所推动的养殖产业空间重构变化研究结果互相印证[37]。其二，工业废气标签数量自2017年起已开始大幅下降，粤港澳大湾区等地的空间异质性同步降低，验证了公众对大气污染防治政策成效的积极反馈[38]。其三，早在2017年，社会生活娱乐噪声标签数量已呈现增长态势，在经济发达、人口密集区域表现出高度的空间异质性。该趋势先于2021年新《中华人民共和国噪声污染防治法》的出台，更早捕捉到公众对宁静生活环境的需求。

4 结论和政策建议

本研究基于大语言模型，挖掘环保举报非结构文本，通过对比污染类型识别的差异，评估其对提升环境管理效率的潜力，得出以下结论：

（1）基于大语言模型的非结构文本挖掘明显提升了环保举报数据的分析效能，其识别准确率显著优于传统结构化标签分析方法，为环保举报数据价值实现提供了更好的技术路径。未来可进一步优化模型在小样本下发现新污染类型的能力，拓展其在环境风险预警中的应用。

（2）非结构文本在表达公众环境关切方面具有独特的信息优势。相比结构化标签，非结构文本有效降低了公众对易感知污染类型的感知偏差，使隐蔽性污染类型得到更真实的呈现。后续研究可深入探讨不同区域、人群在环境感知与关切表达方式的差异，为精细化管理提供依据。

（3）大语言模型强化了管理部门对公众语言的解读能力。本研究从方法学层面验证了将公众作为“环境监督员”的可行性，拓展了基于公众感知数据提升环境管理效能的研究路径，为构建公众参与式的现代环境治理体系提供支撑。

鉴于我国环保举报工作所积累的成功经验，本研究提出“大众环保”的理念——一种基于公众环境信息大数据，通过数据科学手段将非结构的公众关切转化为可供决策参考的有效信息，并以数据融合、集群智慧与公众感知闭环为核心，促进环境治理向源头预防、精细化管理转变的治理理念。其实现路径包括：一是整合数据资源，打通公众关切多源信息采集渠道，构建全域覆盖的公众感知网络，提高环境态势感知能力；二是强化人工智能基座，推动大语言模型等新技术在细分领域的适配，积极探索“人工智能+环保”交叉应用场景，驱动创新技术向治理实践的转化能力；三是构筑公众感知反馈闭环，在环境治理机制设计中更加重视公众诉求快速响应，形成公众感知与政策制定正向反馈，提升环境协同治理水平。

致谢：

感谢生态环境部环境应急与事故调查中心对本研究的支持与指导。

参考文献

[1] 生态环境部 . 关于印发《“十四五”生态环境监测规划》的通知 [EB/OL]. (2021-12-28).[2025-11-04]. https://www.gov.cn/zhengce/zhengceku/2022-01/24/content_5670173.htm.

[2] WU W Q, WANG W W, ZHANG L, et al. Does the public haze pollution concern expressed on online platforms promoted pollution control? – Evidence from Chinese online platforms[J]. Journal of cleaner production, 2021, 318: 128477.

[3] XIE Z H. China’s historical evolution of environmental protection along with the forty years’reform and opening-up[J]. Environmental science and ecotechnology, 2020, 1: 100001.

[4] 生态环境部 . 生态环境部公布 2019 年度全国“12369”环保举报情况 [EB/OL]. (2020-05-13).[2025-04-05]. https://www.mee.gov.cn/xxgk2018/xxgk/xxgk15/202005/t20200513_779050.html.

[5] LI R Q, ZHOU Y C, BI J, et al. Does the Central Environmental Inspection actually work?[J]. Journal of environmental management, 2020, 253: 109602.

[6] DISTELHORST G, HOU Y. Constituency service under nondemocratic rule: evidence from China[J]. The journal of politics, 2017, 79(3): 1024-1040.

[7] HANSEN M H, LI H T, SVARVERUD R. Ecological civilization: interpreting the Chinese past, projecting the global future[J]. Global environmental change, 2018, 53: 195-203.

[8] LI W X, LIU J Y, LI D D. Getting their voices heard: three cases of public participation in environmental protection in China[J]. Journal of environmental management, 2012, 98: 65-72.

[9] DASGUPTA S, WHEELER D. Citizen Complaints as Environmental Indicators: Evidence from China[M]. Washington: The World Bank, 1997.

[10] PENG X, LI Y H, SI Y T, et al. A social sensing approach for everyday urban problem-handling with the 12345-complaint hotline data[J]. Computers, environment and urban systems, 2022, 94: 101790.

[11] SUN Y, JIN F X, ZHENG Y, et al. A new indicator to assess public perception of air pollution based on complaint data[J]. Applied sciences, 2021, 11(4): 1894.

[12] SI Y T, XU L Y, PENG X, et al. Comparative diagnosis of the urban noise problem from infrastructural and social sensing approaches: a case study in Ningbo, China[J]. International journal of environmental research and public health, 2022, 19(5): 2809.

[13] DONG Y L, ISHIKAWA M, LIU X B, et al. The determinants of citizen complaints on environmental pollution: an empirical study from China[J]. Journal of cleaner production, 2011, 19(12): 1306-1314.

[14] KONTOKOSTA C E, HONG B. Bias in smart city governance: how socio-spatial disparities in 311 complaint behavior impact the fairness of data-driven decisions[J]. Sustainable cities and society, 2021, 64: 102503.

[15] SEINEN T M, KORS J A, VAN MULLIGEN E M, et al. Using structured codes and free-text notes to measure information complementarity in electronic health records: feasibility and validation study[J]. Journal of medical internet research, 2025, 27: e66910.

[16] FRAISL D, SEE L, FRITZ S, et al. Leveraging the collaborative power of AI and citizen science for sustainable development[J]. Nature sustainability, 2024, 8(2): 125-132.

[17] LAZER D M J, PENTLAND A, WATTS D J, et al. Computational social science: obstacles and opportunities[J]. Science, 2020, 369(6507): 1060-1062.

[18] GINSBERG J, MOHEBBI M H, PATEL R S, et al. Detecting influenza epidemics using search engine query data[J]. Nature, 2009, 457(7232): 1012-1014.

[19] GONZÁLEZ-BAILÓN S, WANG N, RIVERO A, et al. Assessing the bias in samples of large online networks[J]. Social networks, 2014, 38: 16-27.

[20] BICKERSTAFF K, WALKER G. Public understandings of air pollution: the‘localisation’of environmental risk[J]. Global environmental change, 2001, 11(2): 133-145.

[21] DUNLAP R E, VAN LIERE K D. The “new environmental paradigm”[J]. The journal of environmental education, 1978, 9(4): 10-19.

[22] SLOVIC P. Perception of risk[J]. Science, 1987, 236(4799): 280-285.

[23] ELMAS T. The impact of data persistence bias on social media studies[C]//Proceedings of the 15th ACM Web Science Conference 2023. Austin: ACM, 2023: 196-207.

[24] DU X, KOWALSKI M, VARDE A S, et al. Public opinion matters: mining social media text for environmental management[J]. ACM SIGWEB newsletter, 2020, 2019(Autumn):1-15.

[25] JIANG H C, LIN P, QIANG M S. Public-opinion sentiment analysis for large hydro projects[J]. Journal of construction engineering and management, 2016, 142(2): 05015013.

[26] 方小飞 , 黄孝喜 , 王荣波 , 等 . 基于 LDA 模型的移动投诉文本热点话题识别 [J]. 数据分析与知识发现 , 2017,1(02): 19-27.

[27] LONGADGE R, DONGRE S. Class imbalance problem in data mining review[EB/OL]. (2013-05-08).[2025-11-04]. https://arxiv.org/abs/1305.1707.

[28] ROUMELIOTIS K I, TSELIKAS N D, NASIOPOULOS D K. DeepSeek and GPT fall behind: Claude leads in zero-shot consumer complaints classification[EB/OL]. (2025-02-10)[2025-11-04]. https://www.preprints.org/Manuscript/202502.0720.

[29] GAO T C, JIN J S, KE Z T, et al. A comparison of DeepSeek and other LLMs[EB/OL]. (2025-02-26).[2025-11-04]. https://arxiv.org/abs/2502.03688.

[30] JUNG Y. A semantic annotation framework for scientific publications[J]. Quality & quantity, 2017, 51(3): 1009-1025.

[31] LESKOVEC J, ADAMIC L A, HUBERMAN B A. The dynamics of viral marketing[J]. ACM transactions on the web, 2007, 1(1): 5.

[32] ARAL S, WALKER D. Identifying influential and susceptible members of social networks[J]. Science, 2012, 337(6092): 337-341.

[33] 生态环境部大气环境司 . 关于印发《2018—2020 年全国恶臭 / 异味污染投诉情况分析》报告的函 [EB/OL]. (2021-07-26).[2025-08-26]. https://www.mee.gov.cn/xxgk2018/xxgk/sthjbsh/202108/t20210802_853623.html.

[34] MOYANO E, PANIAGUA A, LAFUENTE R. Environmental policy, public opinion and global climate change in southern Europe: the case of Andalusia[J]. Open environmental sciences, 2008, 2(1): 62-70.

[35] LOPEZ A, GARZA R. Do sensory reviews make more sense? The mediation of objective perception in online review helpfulness[J]. Journal of research in interactive marketing, 2022, 16(3): 438-456.

[36] FRAISL D, HAGER G, BEDESSEM B, et al. Citizen science in environmental and ecological sciences[J]. Nature reviews methods primers, 2022, 2(1): 64.

[37] 周建军 , 谭莹 , 胡洪涛 . 环境规制对中国生猪养殖生产布局与产业转移的影响分析 [J]. 农业现代化研究 , 2018, 39(3): 440-450.

[38] 范可欣 , 高健 , 陕亮 , 等 . 工业源大气污染物排放、检测和监管现状及展望 [J]. 环境工程技术学报 , 2025, 15(1): 193-202.

Research on Improving Management Efficiency of Environmental Complaints Based on Large Language Model

WANG Ximing[1,2], ZHONG Lijin[3], DING Longzhen[1,2], LYU Guangfeng[3], QI Xingyu[3], HU Qing[1,2*]

(1. State Key Laboratory of Soil Pollution Control and Safety, Southern University of Science and Technology, Shenzhen 518055, China; 2. School of Environmental Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China; 3. Beijing Huanding Environmental Big Data Institute, Beijing 100083, China)

Abstract: Environmental complaints provide a vital channel for public participation in environmental governance, through which the public provides a large volume of unstructured textual data. However, traditional methods relying on structured labels cannot effectively process this information, limiting the capacity to identify potential environmental issues from these complaints. Based on 567,985 national environmental complaint records from 2016 to 2021, this study introduces the DeepSeek large language model to mine unstructured text and systematically compares its performance with structured labels in identifying pollution types. The findings reveal that: ①Unstructured text improves the accuracy of pollution-type identification, achieving a higher median F1-score (0.92) compared to structured labels (0.50); ②Unstructured text reduces perceptual bias toward easily perceived pollution types, with a 9.6% decrease in association rule confidence difference, thereby more accurately reflecting the actual environmental concerns of the public;③Unstructured text is more sensitive to spatiotemporal features, enabling early detection of shifts in public concerns. Case studies indicate that this method advanced the identification of aquaculture wastewater and industrial exhaust gas by approximately four and two years, respectively, and revealed a growing public focus on noise pollution. This study establishes a novel LLM-based approach for rapidly identifying latent environmental concerns from unstructured text, supporting regulators in swiftly responding to public needs and promoting refined governance.

Keywords: environmental complaint; large language model; text mining; public participation; citizen science; environmental management