

基于序关系与聚类分析的气田站场关键放空排放源识别

苏诗漫¹, 吴曼琪², 王琳玲¹, 李祎涵¹, 史春艳¹, 伍颖^{2*}

¹中国石油天然气股份有限公司西南油气田分公司川东北气矿, 四川 达州

²西南石油大学土木工程与测绘学院, 四川 成都

收稿日期: 2023年11月29日; 录用日期: 2024年1月23日; 发布日期: 2024年2月28日

摘要

放空排放是气田生产中重要环节之一, 双碳背景下具体放空排放量的获取及关键排放源识别一直是实现气田生产放空减排的重要途径。鉴于气田站场放空数据的复杂性及其聚合性、不完整性及其数据量庞大等特征, 本文提出了基于序关系法与聚类分析的关键排放源识别模型。首先依据气田站场工艺流程特征构建了4类30种生产中放空排放量化体系; 其次采用序关系法并结合站场实际计算得到站场内具体放空排放源的排放量; 再应用逆距离插值法对缺失值进行补充, 采用三种聚类分析算法对关键排放源进行识别, 对比分析结果并综合各算法特点, 找到更加平衡的分类阈值并获取分类结果; 最后将该方法应用于西南某气田, 分析得出新的分类阈值为2.7756, 并发现三种算法的关键排放源均主要分布在增压站中。研究表明, 识别结果合理可靠, 为关键放空排放源识别工作提供了新思路。

关键词

温室气体减排, 气田站场, 关键排放源, 序关系法, 聚类分析

Identification of Key Venting Emission Sources at Gas Field Stations Based on Ordinal Relationship Method and Cluster Analysis

Shiman Su¹, Manqi Wu², Linling Wang¹, Yihan Li¹, Chunyan Shi¹, Ying Wu^{2*}

¹PetroChina Southwest Oil & Gasfield Company Northeast Sichuan Gas District, Dazhou Sichuan

²School of Civil Engineering and Geomatics, Southwest Petroleum University, Chengdu Sichuan

Received: Nov. 29th, 2023; accepted: Jan. 23rd, 2024; published: Feb. 28th, 2024

*通讯作者。

文章引用: 苏诗漫, 吴曼琪, 王琳玲, 李祎涵, 史春艳, 伍颖. 基于序关系与聚类分析的气田站场关键放空排放源识别[J]. 环境保护前沿, 2024, 14(1): 121-135. DOI: 10.12677/aep.2024.141018

Abstract

Air release emission is one of the important links in gas field production, and the acquisition of specific air release emission and identification of key emission sources in the context of dual-carbon has been an important way to realize air release emission reduction in gas field production. In view of the complexity, aggregation, incompleteness and huge data volume of air release data from gas field stations, this paper proposes a key emission source identification method based on the ordinal relationship method and cluster analysis. First of all, based on the characteristics of gas field station process flow, we constructed a system of 4 categories and 30 types of air release emissions in production; secondly, we used the ordinal relationship method and combined with the actual calculation of the station to get the emissions of specific air release emission sources; then we applied the inverse distance interpolation method to supplement the missing values, and used three kinds of cluster analysis algorithms to identify the key emission sources, compared the results of the analysis and synthesized the characteristics of the algorithms, and then we found the balanced classification thresholds and obtained the classification results. Finally, the method is applied to an actual case in a gas field in Southwest China, and the new classification threshold is 2.775581, and it is found that the key emission sources of the three algorithms are mainly distributed in the booster station. The study shows that the identification results are reasonable and reliable, and provides new ideas for the identification of key venting emission sources.

Keywords

Greenhouse Gas Emission Reduction, Gas Mine Production, Key Emission Sources, Ordinal Relationship Method, Cluster Analysis

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

党的二十大报告指出：“积极参与应对气候变化全球治理”，同时强调“积极稳妥推进碳达峰碳中和”。随着中国经济的高速发展，油气行业也在飞速的发展，在油气行业生产过程中不可避免的会产生大量的甲烷排放，造成全球范围内甲烷排放量日益增多。甲烷是一种短期高效的温室气体，油气行业是中国主要的甲烷排放源之一，减少油气行业的甲烷排放对于保护大气环境，实现我国 2060 碳中和的目标有重要的现实意义[1]-[8]。研究油气行业甲烷减排是油气领域研究的热点之一。

目前，国内外相关学者对油气行业的甲烷排放领域进行了大量研究。张仁健等[9]结合油气活动水平及单一的甲烷排放因子，计算得出 1990 年和 1994 年我国油气行业甲烷逃逸排放量；Zhang 等[10]估算了我国 1980~2007 年能源活动相关的甲烷排放，采用具体的排放因子，识别了我国油气行业的甲烷逃逸排放及相应变化趋势。Peng 等[11]参考国外文献的相关排放因子，核算了 1980~2010 年我国主要人为源的甲烷排放情况，分析了油气行业甲烷逃逸排放的空间分布特点。Michael Mac Kinnon [12]介绍了一种新的基于原因的方法来评估每单位燃料消耗的边际甲烷排放量，采用宏观的方法将甲烷排放源分为基于时间的、基于随机事件的(气动控制器)和基于吞吐量的排放三大类。Qining Chen 等[13]引入生命周期法，提出排放清单整合分配法和分解分配法对 Eagle Ford 盆地甲烷排放进行核算。黄满堂等[14]考虑了 IPCC (联合国政府间气候变化专门委员会)排放因子缺省值的不足，对 2015 年我国油气行业的甲烷逃逸排放进行了测算，相关结果涉

及省级尺度甲烷排放的具体情况。目前针对油气行业甲烷减排的研究存在一些不足：一是在对油气行业的甲烷排放进行研究时，研究者通常将放空排放作为其中的一部分考虑在内；二是放空排放作为气田生产中重要的一环，但针对放空排放的研究主要集中在系统的优化[15]，对于放空排放的量化缺少针对性的研究。对气田站场的甲烷放空排放进行量化时，目前最精确的方法是通过实测设备获取排放数据[16][17]。然而，实测困难可能导致天然气站场的甲烷放空排放数据收集出现未记录单个设备的排放量，仅有总排放量的数据及数据缺失等问题，将限制对于甲烷放空排放的深入分析和准确评估。此外，由于气田站场所涉及的工艺复杂性，其数据量庞大，难以透过传统数据处理方法进行高效处理和分析。在这种情况下，针对未标记的大规模数据，无监督机器学习展现出更高效、更精确、更快挖掘数据特性的优势。综上可见，当前学者针对甲烷放空减排中的量化数据对实测技术的依赖性较强，针对具体放空排放研究还不够全面，关键排放源的人工识别难度大，在气田场站生产实际运用中具有一定的局限性。

鉴于此，考虑到实测困难导致的数据聚合性、不完整性及其数据量庞大等特征，本文提出了基于序关系法与聚类分析的关键排放源识别方法。首先采用序关系法计算专家决策下的原始指标权重，结合站场实际情况将原始指标权重归一化进而获取站场内具体放空排放源排放量，以解决排放数据聚合性问题；同时为进一步解决数据不完整性问题并提高关键排放源识别结果精确度，应用逆距离插值法对缺失值进行了预处理，再基于处理后的数据采用 DBSCAN (基于密度的空间聚类算法)、KMEANS (K 均值聚类算法)、层次聚类三种聚类分析算法分别对关键排放源进行识别并将分析结果进行对比，综合三种算法的特点找到一个更加平衡的分类阈值，以达到更准确和可靠的分类结果；最后以西南地区某气田为研究对象，进行关键放空排放源识别，验证识别方法的适用性，以期气田关键放空排放源识别提供一种科学可靠的识别方法。

2. 基于序关系与聚类分析的关键放空排放源识别模型

本文提出基于序关系与聚类分析的关键放空排放源识别模型。模型具体的应用过程如下：① 依据气田站场工艺流程特征，筛选提取站场典型放空排放操作，构建气田站场放空操作量化指标体系；② 基于序关系，采用专家打分法获取相邻指标之间的相对重要度，计算得到原始指标权重，经过归一化处理得到最终的指标得分；③ 根据某月某放空操作排放量计算公式计算本站场内所有放空排放源排放量，并应用逆距离权重插值法补充缺失的异常值；④ 通过算法编程，导入数据进行聚类分析，获取分析结果，并对比三种不同算法结果，最终综合各算法特点，找到更加平衡的分类阈值，从而识别出所有场站中所包含的具体关键排放源，并可依据识别结果针对性的提出管控措施。基于序关系与聚类分析的关键放空排放源识别模型实现流程见图 1。

序关系法的运用为解决排放数据聚合性问题提供了基础。无监督机器学习算法中的聚类分析被用于对关键排放源的识别，其中，为填补数据中的缺失值，采用逆距离插值法预处理放空气量数据，从而解决数据不完整性的问题；在数据预处理的基础上分别采用 DBSCAN 算法、KMEANS 算法、层次聚类算法对关键排放源进行识别。

2.1. 气田站场量化指标体系构建

气田站场内的工艺流程因应用的技术、生产阶段和具体操作的不同而呈现多样性、复杂性、交叉性。在指标选取时，总体上应该遵循科学性、系统性、全面性、典型性、可测算性等原则。不同的生产阶段涉及到的放空操作具有差异性，这包括对不同设备、管道系统或处理站点的放空处理。为了更好地管理和监控这种多样性，有必要对站场进行分类，以便对每个类别选取相应的放空操作量化指标。依据场站生产过程中对应的角色特征将站场分为了采气站、回注站、集气站、增压站、输配气站 5 类，在考虑到回注站中放空操作较少的情况下，这一类站点并未被纳入综合考虑范畴。综合考虑 4 类站场工艺流程特

点，分类归纳了 30 个指标，并征求了现场工作人员、现场管理人员以及拥有丰富气田现场经验专家的意见，最终构建了气田站场放空操作量化指标体系，见图 2。

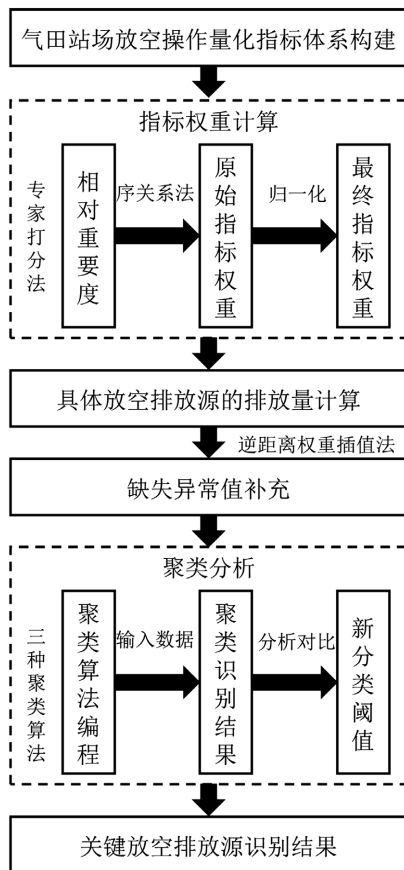


Figure 1. Flowchart of the realization of the method for identifying key air release emission sources based on the ordinal relationship method and cluster analysis

图 1. 基于序关系与聚类分析的关键放空排放源识别方法实现流程图

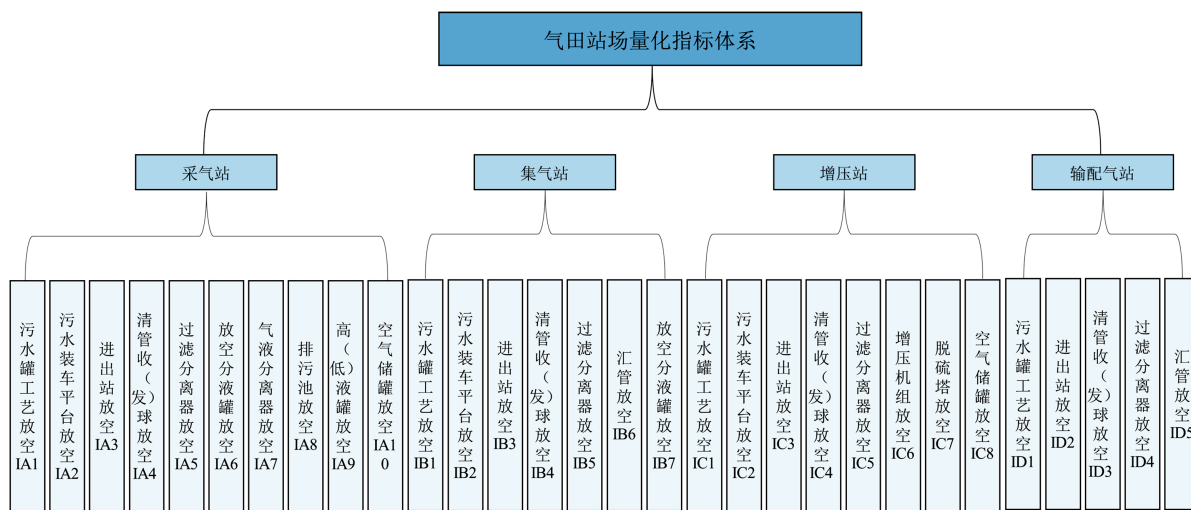


Figure 2. Quantitative index system for gas field station venting operation

图 2. 气田站场放空操作量化指标体系

2.2. 权重指标计算

序关系法[18] [19]不需要构建判断矩阵,且无一致性检查,可操作性强。由于考虑两个指标的共同作用,减少主观分析的偏差,不区分重要度可用于确定气田生产中甲烷放空排放操作排放量体的权重。序关系法确定权重的步骤如下:

(1) 确定指标之间的序关系

根据研究对象中各项评价指标的重要度,确定各指标之间的排序关系。若评价指标 X_i 相对于指标 X_j 重要,则记为 $X_i > X_j$,根据这一原则,结合专家给的评定建议,确定评价对象中各个指标的重要性排序,将最终得到的 m 个指标的排序定义为 $X_1 > X_2 > \dots > X_k > \dots > X_m$ 。

(2) 给出相邻指标之间的相对重要度的比较判断准则:设专家关于评价对象的评价指标 X_k 和 X_{k-1} 的相对重要程度之比进行赋值:

$$r_k = \frac{w_{k-1}}{w_k} (k = m, m-1, \dots, 2) \quad (1)$$

$$r_{k-1} \geq \frac{1}{r_k} (k = m, m-1, \dots, 3) \quad (2)$$

式中: r_k 、 r_{k-1} ——相对重要度;

w_k 、 w_{k-1} ——指标权重;

r_{k-1} 的赋值可参考表 1。

Table 1. Criteria for relative ranking

表 1. 相对等级划分标准

r_k	说明
1.0	指标 X_{k-1} 与指标 X_k 具有同样的重要性
1.2	指标 X_{k-1} 比指标 X_k 稍微重要
1.4	指标 X_{k-1} 比指标 X_k 明显重要
1.6	指标 X_{k-1} 比指标 X_k 强烈重要
1.8	指标 X_{k-1} 比指标 X_k 极端重要

(3) 指标权重计算

专家给出了 r_k 理性赋值,则如式(3)、(4),权重 w_k 为:

$$w_k = \left(1 + \sum_{k=2}^m \prod_{i=k}^m r_i \right)^{-1} \quad (3)$$

$$w_{k-1} = r_k w_k (k = m, m-1, \dots, 2) \quad (4)$$

式中: r_k 、 r_{k-1} ——相对重要度;

w_k 、 w_{k-1} ——指标权重。

在权重计算中,采用了按照不同类型站场进行计算的方法。具体而言,选择了大部分同类型站场中存在的放空操作进行权重计算。然而,由于某些站场可能不包含某些操作,可能导致某些站场中已有操作权重之和不等于 1 的情况。解决该问题,需要进行归一化操作。

通过归一化，可以确保权重计算结果的准确性和一致性。可以更准确地评估站场中放空操作的重要性和影响程度，为后续的决策制定和资源分配提供了有价值的参考依据。归一化公式如下所示：

$$w_{k \text{ 归一化}} = \frac{w_{k \text{ 原}}}{\sum w} \quad (5)$$

式中： $w_{k \text{ 归一化}}$ ——经过归一化处理后的最终指标权重；

$w_{k \text{ 原}}$ ——采用序关系法获取的原始指标权重。

某月某放空操作排放量计算公式如(6)所示：

$$q_i = w_{k \text{ 归一化}} \cdot Q_i \quad (6)$$

式中： q_i ——某月某放空操作排放量；

Q_i ——某月总放空排放量。

2.3. 缺失异常值补充

数据预处理是数据挖掘和机器学习中不可或缺的环节，它通过降低噪声影响、提高数据一致性和完整性、优化模型性能和泛化能力，以及提高计算效率，可为后续的分析 and 建模提供可靠的基础。

由于获取的站场内放空排放数据具有随机性，并且受时间变化的影响较大，为对数据进行有效的预处理，选择采用逆距离权重插值[20]方法。

$$Z = \sum_{i=1}^m \lambda_i Z_i \quad (7)$$

$$\lambda_i = \frac{d_i^{-n}}{\sum_{i=1}^m d_i^{-n}} \quad (8)$$

$$d_i = \sqrt{(x_i - x)^2 + (y_i - y)^2} \quad (9)$$

式中： Z ——待插点的估算值；

Z_i ——第 i 个样本点的实测值；

d_i ——第 i 个样本点与待插点之间的距离；

m ——参与计算的实测样本点个数；

n ——幂指数，它控制着权重系数随待插点与样本点之间距离的增加而下降的程度， n 越大时，较近的样本点赋予更高的权重， n 越小时，权重更均匀地分配给各样本点。

2.4. 关键排放源识别

聚类分析是一种无监督学习方法，有助于发现数据中的模式、结构和关联关系。进行聚类分析之前，对数据采用逆距离权重插值法对缺失值进行补充，旨在清洁、转换和准备数据以便于后续的聚类分析。DBSCAN 聚类算法基于密度，能够自动发现不同形状和大小的簇，对噪声和异常点具有较好的鲁棒性。但对高维数据和密度变化大的数据效果较差，计算复杂度高。KMEANS 聚类算法简单易懂，适用于维度较低的数据集，能够处理非线性数据且计算速度较快。但需要预先指定聚类数目 k ，对不规则形状的簇效果欠佳，结果受数据初始化影响。层次聚类算法能够发现数据的聚类层次结构，对于不规则形状的簇具有良好表现，但计算复杂度高。通过使用这三种算法，可以根据不同数据特征和需求提供灵活、高效且全面的聚类结果，应对各种数据集的聚类需求。

① DBSCAN 算法

DBSCAN 算法[21]的核心指标是邻域范围半径和最小邻域点数目阈值, 即 ϵ 和 $Minpts$ 。

其中: 邻域范围半径是以某一中心点 p 为圆心、以半径 ϵ 作圆, 圆内范围即为 EPS 邻域, 该邻域中数据点的个数为 m , 可体现该簇的密度大小, 最小邻域点数目值 $Minpts$ 指最少包含数据点的个数, 一般直接给定, 见图 3。

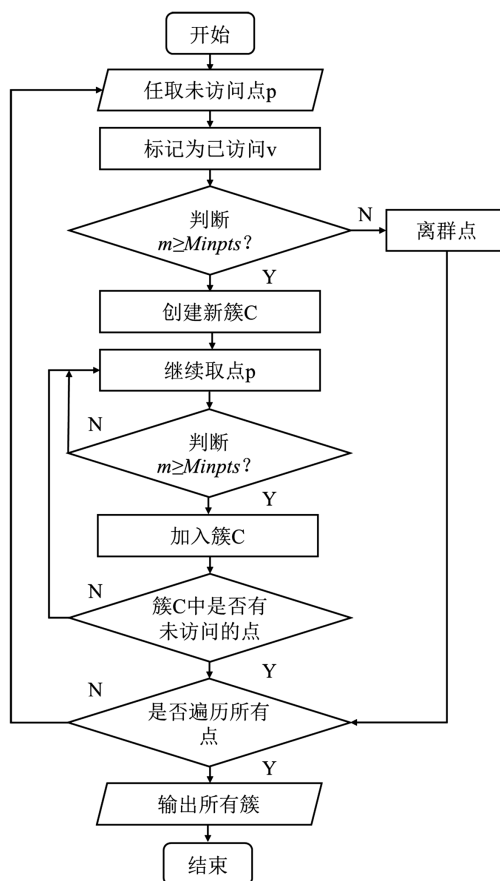


Figure 3. Basic flowchart of DBSCAN algorithm

图 3. DBSCAN 算法基本流程图

DBSCAN 算法的基本流程如下:

- 1) 初始化参数: 设置邻域范围半径(ϵ)和最小邻域点数目($Minpts$);
- 2) 选择起始点: 随机选择数据集中的一个未被访问的数据点 p , 并标记为已访问 v ;
- 3) 判断核心点: 计算以点 p 为中心、半径为 ϵ 的邻域内的数据点数目 m 。如果 m 大于等于 $Minpts$, 则将点 p 标记为核心点并创建新簇 C ;
- 4) 扩展簇: 如果点 p 被标记为核心点, 从点 p 的邻域中选择一个未被访问的数据点 q , 将 q 加入当前簇, 并递归地对点 q 进行邻域扩展, 即继续检查以 q 为中心、半径为 ϵ 的邻域;
- 5) 标记为离群点: 如果点 q 不是核心点, 但仍在核心点的 ϵ -邻域内, 将点 q 标记为离群点;
- 6) 找到下一个未访问的点: 重复步骤 2)~5), 直到所有的数据点都被访问过;
- 7) 输出所有簇: 通过不断扩展簇, 最终形成一组聚类, 其中包括核心点及其可达的所有数据点, 而不可达的点被视为噪声点。

② KMEANS 算法

KMEANS 算法[22]属于基于划分的聚类算法，通过迭代计算来确定样本的簇标签。该算法随机选取 k 个初始聚类中心，每一个聚类中心代表一类数据集簇；然后将数据点划分到最近的数据簇；重新计算数据簇中心点，直到聚类准则函数收敛，收敛函数如式(10)。

$$E = \sum_{i=1}^k \sum_{x \in C_i} \|x - u_i\|_2^2 \tag{10}$$

式中： E ——KMEANS 算法针对样本 $D = \{x_1, x_2, \dots, x_k\}$ 聚类所得簇 $C = \{C_1, C_2, \dots, C_k\}$ 划分的最小化平方误差， u_i 表示为

$$u_i = \frac{1}{|C_i|} \sum_{x \in C_i} x \tag{11}$$

式中： u_i ——簇 C_i 的均值向量。

③ 层次聚类算法

层次聚类算法[23]通过构建树状结构来表示数据的层次关系。该算法从单个样本开始，然后逐渐合并相似的样本或簇，直到形成一个完整的层次聚类结构。它可以帮理解数据的聚类结构，并且可以根据需要选择不同的层次进行聚类，见图 4。

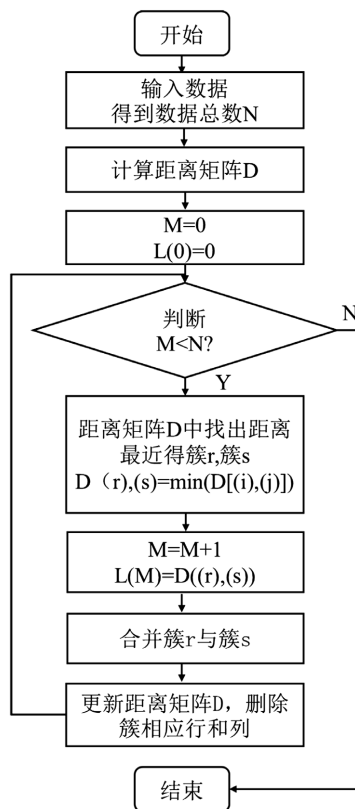


Figure 4. Basic flowchart of hierarchical clustering algorithm
图 4. 层次聚类算法基本流程图

层次聚类算法的基本流程如下：

- 1) 输入：得到数据总数 N ；

- 2) 计算距离矩阵 D : 对于 N 个数据点, 计算所有样本之间的相似度或距离, 得到距离矩阵 D ;
- 3) 初始化: 将每个样本视为一个初始簇, 设置 $M = 0$, $L(0) = 0$;
- 4) 判断 $M < N$: 如果 M 小于 N , 执行以下步骤;
- 5) 找到最近的簇: 从距离矩阵 D 中找到距离最近的两个簇 r 和 s ;
- 6) 合并簇: 合并簇 r 与簇 s ;
- 7) 更新距离矩阵 D : 删除距离矩阵 D 中与簇 r 和簇 s 相关的行和列, 并添加新的行和列, 表示合并后的簇;
- 8) 增加合并次数: $M = M + 1$;
- 9) 更新合并记录: $L(M) = D(I)(s)$;
- 10) 重复步骤④: 回到步骤④, 继续执行, 直到 $M \geq N$ 只剩下一个簇为止, 则聚类过程结束, 得到层次聚类结果。

3. 案例分析

以西南某气田站场为研究背景, 验证本文所提出的基于序关系与聚类分析的关键放空排放源识别方法在气田生产中关键放空排放源识别方面的适用性。该气田有五种不同类型的站场, 共计 89 个。站场类型包括采气站、回注站、集气站、增压站和输配气站。其中, 采气站的数量为 52 个, 回注站的数量为 7 个, 集气站的数量为 14 个, 增压站的数量为 11 个, 输配气站的数量为 5 个。

通过现场调研, 获取了 2022 年 1 月至 2023 年 6 月的《环保月报数据统计表》。该表提供了不同站场的总放空排放数据, 从而允许采用序关系法对不同类型站场中具体设备的放空排放量进行计算和分析。通过对这些数据的整合和研究, 可以更好地了解甲烷放空排放状况, 并为制定相应的管理策略和措施提供可靠的依据。

经过对大量放空排放量数据的梳理后, 发现数据量庞大, 人工进行关键排放源的识别工作较为困难。但相比之下, 机器学习的优越性使其成为处理复杂问题和大规模数据的有效工具, 有助于提高工作效率和准确性。

3.1. 指标权重结果

本小节利用序关系法, 通过专家打分法获取相邻指标之间的相对重要度, 该过程可以帮助计算得到原始指标的权重。权重经过归一化处理, 最终得到每个指标的得分。能够有效地衡量指标之间的相对重要性, 为后续的数据分析和处理提供了权衡因素, 进而能够更准确地识别和评估关键的站场放空排放源。

通过实地考察、调研气田站场放空排放数据与站场工艺流程图, 基于实际放空操作。采用序关系法处理专家评分, 根据式(1)~式(4)计算得到气田生产中放空操作指标体系及原始权重见表 2。

Table 2. Weighting of operational indicators for release of airspace at each type of station
表 2. 各类站场放空操作指标权重

站场类型	放空操作指标	指标权重
采气站	污水罐工艺放空 IA1	0.097
	污水装车平台放空 IA2	0.048
	进出站放空 IA3	0.191

续表

采气站	清管收(发)球放空 IA4	0.069
	过滤分离器放空 IA5	0.146
	放空分液罐放空 IA6	0.096
	气液分离器放空 IA7	0.146
	排污池放空 IA8	0.057
	高(低)液罐放空 IA9	0.085
	空气储罐放空 IA10	0.064
集气站	污水罐工艺放空 IB1	0.096
	污水装车平台放空 IB2	0.070
	进出站放空 IB3	0.234
	清管收(发)球放空 IB4	0.101
	过滤分离器放空 IB5	0.176
	汇管放空 IB6	0.185
增压站	放空分液罐放空 IB7	0.137
	污水罐工艺放空 IC1	0.091
	污水装车平台放空 IC2	0.070
	进出站放空 IC3	0.221
	清管收(发)球放空 IC4	0.106
	过滤分离器放空 IC5	0.151
	增压机组放空 IC6	0.175
	脱硫塔放空 IC7	0.119
空气储罐放空 IC8	0.067	
输配气站	污水罐工艺放空 ID1	0.134
	进出站放空 ID2	0.285
	清管收(发)球放空 ID3	0.144
	过滤分离器放空 ID4	0.213
	汇管放空 ID5	0.223

以四川某气田某井 2022 年 1 月~2023 年 6 月放空数据为例计算归一化后指标权重, 该井包含的放空操作与放空指标相比缺少了污水装车平台放空 IA2、放空分液罐放空 IA6、气液分离器放空 IA7、排污池放空 IA8、高(低)液罐放空 IA9、空气储罐放空 IA10 六项, 根据式(5)计算得到归一化后计算权重见表 3。

Table 3. Calculated weights after normalization for a well
表 3. 某井归一化后指标权重

站场名称	放空操作	原始计算权重	归一化计算权重
某井	污水罐工艺放空 IA1	0.097	0.192
	进出站放空 IA3	0.191	0.380
	清管收(发)球放空 IA4	0.069	0.137
	过滤分离器放空 IA5	0.146	0.291
	其他	0	0
	合计	0.503	1

3.2. 数据预处理分析

根据式(6)计算得出样本站场中所有具体放空排放源的排放量。以 2023 年 1 月至 2023 年 6 月期间为例, 某井放空数据未经过数据预处理的情况下, 发现存在 5 项数据均为 0 且与前几月数据相差太大的异常情况, 这些数据可以被认为是需要处理的异常值。为解决这个问题, 采用逆距离权重插值法和归一化方法对原始数据进行预处理, 以补充缺失的异常值。根据式(6)计算得到每月具体放空排放源排放量及数据预处理前后对比, 分别见图 5 和图 6。

通过逆距离权重插值法和归一化方法对样本站场中所有具体放空排放源的原始数据进行预处理后, 经过仔细的分析与对比, 发现增压站相对于其他站场表现出更高的排放浓度和排放量。这一结果揭示了增压站在放空排放方面的显著特征, 可为后续的关键排放源识别工作提供有力的参考依据。

3.3. 关键排放源识别结果及分析

基于数据梳理结果, 该气田一共有 90 种具体的放空操作, 将不同站场中具体的放空操作对应固定组数编号方便进行聚类分析(1~29 为采气站的放空操作, 30~44 为集气站的放空操作, 45~90 为增压站的放空操作, 见图 7~10)。依据 1.2 中阐述的三种聚类分析算法思想分别编写出三类聚类分析算法的代码, 将梳理并预处理的具体放空排放源的排放量数据导入不同代码中进行聚类分析, 根据放空排放量的差异将数据划分为两个类别, 其中 1 类定义为放空排放量低, 2 类定义为放空排放量大。

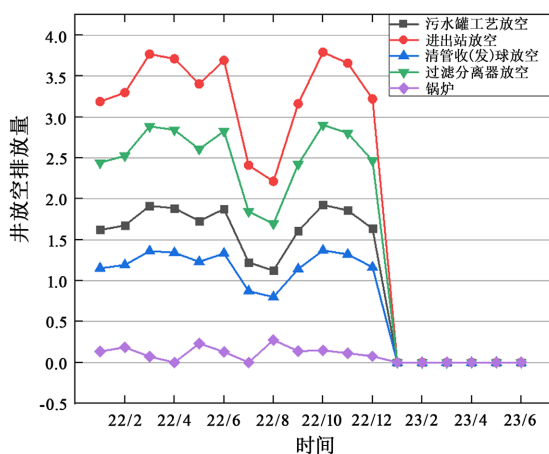


Figure 5. Basic flowchart of hierarchical clustering algorithm

图 5. 层次聚类算法基本流程图

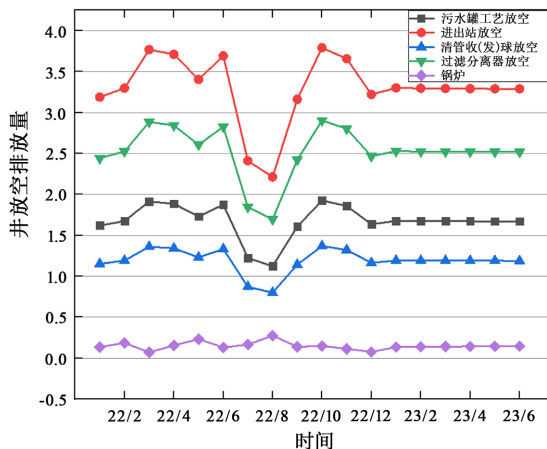


Figure 6. A well venting data after preprocessing
图 6. 某井放空数据预处理后

采用这三种典型的聚类分析算法(DBSCAN 算法、KMEANS 算法、层次聚类算法)可以得到较为全面的聚类分析结果,进而可以充分挖掘数据中的内在关系和模式。三种算法是比较常见的算法,并且能够依据数据特征进行聚类分析,但是需要明确每种算法的特点:第一种 DBSCAN 算法不需要事先知道簇的个数,因此对于不确定簇数量的情况更为适用,且分类阈值较为冒险;第二种 KMEANS 算法需要用户事先指定簇的个数,对于数据结构较为清晰、聚类数目已知的情况比较适用,且分类阈值较为保守;第三种层次聚类算法适用于数据具有明显的层次结构,可用于揭示数据的组织关系,无需事先知道簇的个数,能够在不同层次上提供聚类结果,且分类阈值较为适中。为了平衡各算法的优劣,故综合三种算法特点,取三种算法分类阈值的均值作为新分类阈值进行第四种聚类分析。接下来将对这四种聚类分析结果进行进一步分析,深入理解数据并获得更准确的聚类结构。基于识别关键排放源的目的,需要根据四种不同聚类分析算法得出三项结果:分类阈值、放空排放量低的排放源数量及放空排放量高的排放源数量。四种聚类分析的分类阈值及结果见表 4 及图 7~10。

(1) 聚类分析结果

综合四种算法的特点与表 4 及图 7~10 的聚类分析结果进行分析,结果如下:第一种 DBSCAN 算法较为冒险,选取的分类阈值为 1.2758;聚类分析的结果显示:1 类有 54 个,2 类有 36 个,关键排放源主要集中在增压站中;第二种 KMEANS 算法较为保守,选取的分类阈值为 4.36578;聚类分析的结果显示:1 类有 76 个,2 类有 14 个,关键排放源主要集中在增压站中;第三种层次聚类算法较为适中,选取的分类阈值为 2.6853;层次聚类分析的结果显示:1 类有 68 个,2 类有 22 个,关键排放源主要集中在增压站中;三个阈值的均值为 2.7756,均值聚类分析的结果显示:1 类有 68 个,2 类有 22 个。

(2) 分类阈值确定

基于上述分析结果进一步确定最终使用的分类阈值。首先对分类阈值进行比较:KMEANS > 均值 > 层次聚类 > DBSCAN,这是因为每种算法在处理数据时采用了不同的计算方法和假设。接着对 2 类(关键排放源)分类个数进行比较:DBSCAN > 均值 = 层次聚类 > KMEANS。针对上述排序分析可知:DBSCAN 算法选取的分类阈值较为冒险,可能导致较多样本被误分类为噪声或者未能正确聚类;KMEANS 算法选取的分类阈值较为保守,可能将一些相似的样本分配到不同的类别中;层次聚类算法的分类阈值选取较为适中,能够较好地将样本进行分类,且与实际情况更为符合;综合考虑选取三个分类阈值的均值 2.7756,均值聚类分析的结果与层次聚类算法相同,表明新的分类阈值能够较好地地区分出样本的类别。

(3) 关键排放源识别结果

采用最终选定的均值 2.7756 作为分类阈值，对 90 种具体放空操作进行聚类分析。聚类分析的结果如下：① 采气站中的关键排放源为进出站放空与过滤分离器放空，增压站中的关键排放源集中在进出站放空、清管收(发)球放空、过滤分离器放空、增压机组放空；② 增压站表现出更高的排放浓度和排放量，说明了增压站是一个潜在的非常重要的放空排放源，并且也提示了增压站的排放可能对环境造成更大的影响，需要优先加强管理；其次，尽管通过三种聚类算法得到的分类结果有一定差异，但总的趋势表明增压站是关键排放源的主要分布区域，也证明了该结论的合理性。

Table 4. List of classification thresholds and results

表 4. 分类阈值及结果一览表

算法名称	分类阈值	放空排放量低	放空排放量高
DBSCAN	1.2758	54	36
KMEANS	4.3657	76	14
层次聚类	2.6853	68	22
均值	2.7756	68	22

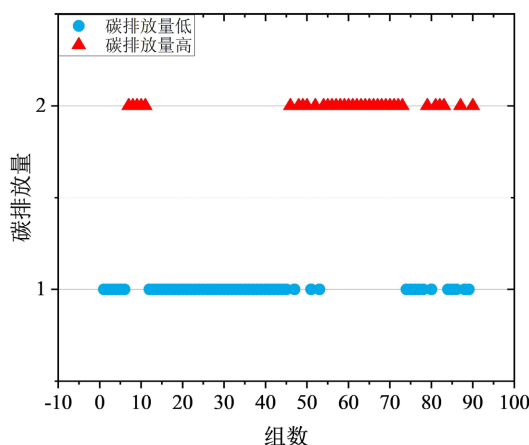


Figure 7. DBSCAN algorithm

图 7. DBSCAN 算法

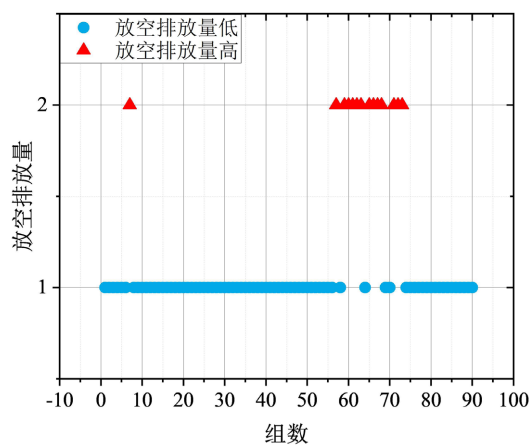


Figure 8. KMEANS algorithm

图 8. KMEANS 算法

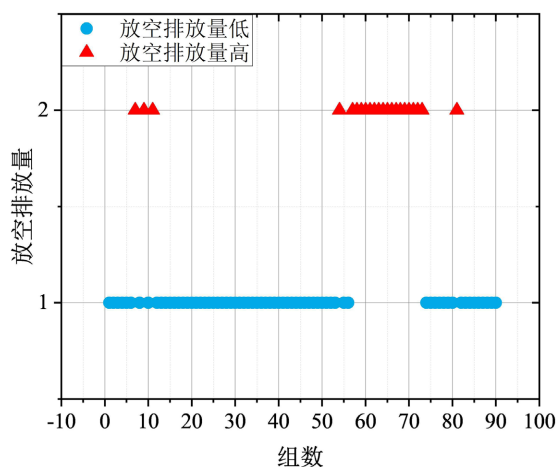


Figure 9. Hierarchical clustering algorithm

图 9. 层次聚类算法

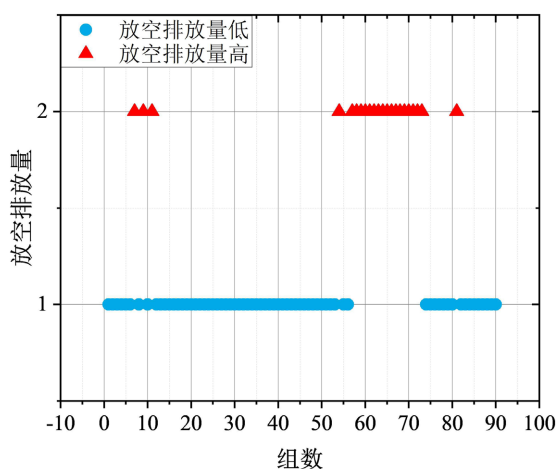


Figure 10. Mean value

图 10. 均值

4. 结论

① 针对气田放空排放实测困难导致的数据聚合性、不完整性及其数据量庞大等特征，提出了一套基于序关系与聚类分析的关键放空排放源识别模型，该模型首先采用序关系法计算专家决策下的原始指标权重，结合站场实际情况将原始指标权重归一化获取了站场内具体放空排放源排放量，解决了排放数据聚合性问题，提高了模型的综合性和适应性；同时，应用逆距离插值法对缺失值进行了预处理，进一步解决了数据不完整性问题并提高了关键排放源识别结果精确度；再基于处理后的数据采用 DBSCAN（基于密度的空间聚类算法）、KMEANS（K 均值聚类算法）、层次聚类三种聚类分析算法分别对关键排放源进行识别并将分析结果进行对比，综合三种算法的特点找到一个更加平衡的分类阈值，达到了更准确和可靠的分类结果，该模型的综合处理方法可以为类似环境中的排放源识别提供参考，并在实际应用中帮助决策者更准确地了解关键排放源的特征和量化信息，从而更有效地进行环境管理和控制；

② 将该方法应用于某气田，计算分析得到聚类分析的新阈值 2.7756，基于该阈值识别出关键排放源主要分布在增压站的进出站放空、清管收(发)球放空、过滤分离器放空、增压机组放空，未来可依据关键放空排放源的识别结果采取有针对性的减排措施。

参考文献

- [1] World Meteorological Organization (WMO) (2020) WMO Provisional Report on the State of the Global Climate in 2020. WMO, Geneva.
- [2] 李北陵. 甲烷减排, 既是挑战又是机遇[N]. 中国石化报, 2021-11-26(005).
- [3] 张道伟. 四川盆地未来十年天然气工业发展展望[J]. 天然气工业, 2021, 41(8): 34-45.
- [4] 周守为, 朱军龙. 助力“碳达峰、碳中和”战略的路径探索[J]. 天然气工业, 2021, 41(12): 1-8.
- [5] 李鹭光. 中国天然气工业发展回顾与前景展望[J]. 天然气工业, 2021, 41(8): 1-11.
- [6] 汪莉丽, 闫强, 周凤英. 天然气利用有效性的指标体系研究[J]. 可持续发展, 2018, 8(1): 58-64.
<https://doi.org/10.12677/SD.2018.81007>
- [7] 方昆升, 汪彪. 碳中和目标下的美丽中国低碳发展路径研究[J]. 气候变化研究快报, 2023, 12(2): 352-360.
<https://doi.org/10.12677/CCRL.2023.122037>
- [8] 徐瑞. 西南地区天然气勘探开发环境影响评价存在的挑战和探析[J]. 世界生态学, 2023, 12(3): 304-307.
<https://doi.org/10.12677/IJE.2023.123037>
- [9] 张仁健, 王明星, 李晶, 杨昕, 王秀玲. 中国甲烷排放现状[J]. 气候与环境研究, 1999(2): 67-75.
- [10] Zhang, B., Chen, G.Q., Li, J.S. and Tao, L. (2014) Methane Emissions of Energy Activities in China 1980-2007. *Renewable and Sustainable Energy Reviews*, **29**, 11-21. <https://doi.org/10.1016/j.rser.2013.08.060>
- [11] Peng, S.S., Piao, S.L. and Bousquet, P. (2016) Inventory of Anthropogenic Methane Emissions in Mainland China from 1980 to 2010. *Atmospheric Chemistry and Physics*, **16**, 14545-14562. <https://doi.org/10.5194/acp-16-14545-2016>
- [12] Mac Kinnon, M., Heydarzadeh, Z., Doan, Q., Ngo, C. and Reed, J. (2018) Need for a Marginal Methodology in Assessing Natural Gas System Methane Emissions in Response to Incremental Consumption. *Journal of the Air & Waste Management Association*, **68**, 1139-1147. <https://doi.org/10.1080/10962247.2018.1476274>
- [13] Chen, Q., Dunn, J.B. and Allen, D.T. (2019) Aggregation and Allocation of Greenhouse Gas Emissions in Oil and Gas Production: Implications for Life-Cycle Greenhouse Gas Burdens. *ACS Sustainable Chemistry & Engineering*, **7**, 17065-17073. <https://doi.org/10.1021/acssuschemeng.9b03136>
- [14] 黄满堂, 王体健, 赵雄飞, 谢晓栋, 王德羿. 2015 年中国地区大气甲烷排放估计及空间分布[J]. 环境科学学报, 2019, 39(5): 1371-1380.
- [15] 杨梓诚, 高俊莲, 唐旭, 仲冰, 张博. 中国油气行业甲烷逃逸排放核算与时空特征研究[J]. 石油科学通报, 2021, 6(2): 302-314.
- [16] 孙永彪, 张春香, 解东来, 那媛媛, 张鑫. 天然气系统甲烷排放测量与估算研究现状[J]. 油气田地面工程, 2020, 39(10): 30-37.
- [17] 张星雨, 潘季荣, 解东来, 詹巧玲, 张永清. 天然气行业甲烷排放检测与控制——国外研究进展[C]//中国土木工程学会燃气分会. 中国燃气运营与安全研讨会(第十届)暨中国土木工程学会燃气分会 2019 年学术年会论文集(上册). 天津: 煤气与热力杂志社, 2019: 163-168.
- [18] 王学军, 郭亚军, 兰天. 构造一致性判断矩阵的序关系分析法[J]. 东北大学学报, 2006, 27(1): 115-118.
- [19] 王雯悦, 伍颖, 尤潇, 等. 山区输气管道地震风险评价方法研究[J]. 煤气与热力, 2022, 42(10): 7-12+15.
<https://doi.org/10.13608/j.cnki.1000-4416.2022.10.004>
- [20] 薛树强, 杨文龙, 李保金. 反距离加权插值函数性质及最优插值条件[J]. 测绘科学, 2022, 47(10): 1-7+65.
<https://doi.org/10.16251/j.cnki.1009-2307.2022.10.001>
- [21] 高翀. 基于 DBSCAN 的节点仪器桩号自匹配方法设计与验证[J]. 石油物探, 2023, 62(S1): 45-51.
- [22] 林涛, 赵臻. 最近邻优化的 k-means 聚类算法[J]. 计算机科学, 2019, 46(S2): 216-219.
- [23] 朱瑞文, 王雅文, 林欢, 李帅赞, 吴彦芳. 一种基于层次聚类的测试用例集约简方法[J]. 北京邮电大学学报, 2023, 46(4): 9-14. <https://doi.org/10.13190/j.jbupt.2022-080>