

doi: 10.7690/bgzd.2025.10.009

基于谱聚类的畜牧业气象灾害分布特征提取模型

闫 慧, 向 雯, 窦 花, 李建国

(青海省果洛藏族自治州玛沁县气象局, 青海 玛沁 814000)

摘要: 为了解气象灾害分布特征, 实现有效且准确的气象灾害预警, 降低气象灾害给畜牧业造成的损失, 提出一种基于谱聚类的畜牧业气象灾害分布特征提取模型。通过计算影响程度, 确定 11 个气象灾害指标。对气象灾害指标实施标准化、空位补充以及去噪等处理。利用谱聚类算法计算气象灾害特征值, 得出对应的气象灾害风险程度。区域 1、2、5、7 的灾害特征值呈现上升趋势, 风险程度逐年上升; 区域 6、10 的灾害特征值呈现下降趋势, 风险程度逐年下降; 区域 3、4、8、9 的灾害特征值一直维持在较为稳定的水平上, 其中 3、4、8 区域的风险程度一直保持在低度风险水平, 9 区域的风险程度一直保持在中度风险水平。低度风险区域所占面积更大且多集中在西北角, 中度风险区域且多集中在中部, 高度风险区域且多集中在东南地区。结果表明: 所构建模型具有较好的应用效果, 能够有效实现畜牧业气象灾害分布特征提取。

关键词: 谱聚类算法; 畜牧业; 气象灾害; 气象灾害指标; 分布特征提取模型

中图分类号: TP301.6 **文献标志码:** A

A Model for Extracting Distribution Characteristics of Meteorological Disasters in Animal Husbandry Based on Spectral Clustering

Yan Hui, Xiang Wen, Dou Hua, Li Jianguo

(Meteorological Bureau of Maqin County in Guoluo Tibetan Autonomous Prefecture of Qinghai Province, Maqing 814000, China)

Abstract: In order to understand the distribution characteristics of meteorological disasters, realize the effective and accurate early warning of meteorological disasters, and reduce the losses caused by meteorological disasters to animal husbandry, a model for extracting the distribution characteristics of meteorological disasters in animal husbandry based on spectral clustering was proposed. By calculating the degree of impact, 11 meteorological disaster indicators are determined. Standardization, vacancy supplement and noise removal are carried out for meteorological disaster indicators. Spectral clustering algorithm is used to calculate the characteristic value of meteorological disasters, and the corresponding risk degree of meteorological disasters is obtained. The regions 1, 2, 5 and 7 show an upward trend, and the risk degree increases year by year, while the disaster characteristic values of regions 6 and 10 show a downward trend, and the risk degree decreases year by year; The disaster characteristic values of regions 3, 4, 8 and 9 have been maintained at a relatively stable level, of which the risk level of regions 3, 4 and 8 has been maintained at a low risk level, and the risk level of region 9 has been maintained at a moderate risk level. The low risk area occupies a larger area and is mostly concentrated in the northwest corner, the moderate risk area is mostly concentrated in the middle, and the high risk area is mostly concentrated in the southeast. The results show that the model has a good application effect and can effectively extract the distribution characteristics of meteorological disasters in animal husbandry.

Keywords: spectral clustering algorithm; animal husbandry; meteorological disaster; meteorological disaster index; distribution feature extraction model

0 引言

农业作为支柱性产业, 在维持整个社会的生存与发展中发挥了不可替代的作用。通过农业种植和养殖, 为人们提供了日常生活中不可或缺的基础营养物质。畜牧业是农业中的重要组成部分, 通过畜牧业不仅能够为人们提供肉、奶、蛋类等食物, 而且能提供皮毛等原料^[1]。畜牧业的发展情况与气象条件息息相关, 当发生气象灾害时, 会直接影响地

区草场的生长, 导致牲畜大批死亡, 因此气象信息是各级畜牧部门和畜牧饲养者都时刻关注的焦点^[2]。通过有效了解气象信息, 及时发现警情, 实现灾害预警, 以便牧民及时做好防范措施, 降低气象灾害带来的损失。

在气象灾害预警中, 气象灾害分布特征的了解是不可缺少的一环, 透过气象灾害分布特征能够划分灾害风险等级, 方便牧民和当地部门采取针对性

收稿日期: 2024-10-26; 修回日期: 2024-11-22

第一作者: 闫 慧(1990—), 女, 辽宁人。

策略。关于灾害分布特征的相关研究有很多^[3]。例如，刘琰琰等^[4]以旺苍县为研究地区，该研究中以研究区内 16 个监测站 37 年间的气象资料为基础，获取了气象气象灾害指标，然后利用 ArcGIS 进行分析，划分了灾害等级，最后绘制了灾害特征空间分布图，确定危险性区域。王琼等^[5]以沧州市为研究区，分析该区域内 57 年间的气象灾害数据为基础，提取气象灾害判别指标，然后利用层次分析确定了指标的影响权重，最后确定灾害影响程度，完成气象灾害变化特征分析。

结合前人研究经验，在谱聚类算法的基础上，构建一种畜牧业气象灾害分布特征提取模型。通过该模型研究以期对灾害预警提供可靠的分析手段，提高灾害预警的准确性。

1 畜牧业气象灾害分布特征提取研究

了解气象灾害分布特征对于畜牧业发展，以便制定预防策略，降低灾害损失都具有十分重要的意义^[6]。为此，构建一种基于谱聚类的特征提取模型。该模型构建主要分为气象灾害指标确定、指标预处理、气象灾害分布特征提取。

1.1 气象灾害影响指标确定

为了能够及时且方便地收集气象信息，一般会在研究区设置多个气象监测站，以监测随时变化的气象信息^[7]。这些气象信息就是气象灾害分布特征研究的基础。在获取的气象信息中，通过计算指标对气象灾害的影响程度来选取气象灾害描述性指标，计算公式为：

$$A(a_i, y) = \frac{\sum_{i=1}^n (a_i - \bar{a})^2}{B_y \cdot \ln(n)} \quad (1)$$

式中： $A(a_i, y)$ 为第 i 个指标 a_i 对气象灾害 y 的影响程度； B_y 为气象灾害 y 的信息熵； a_i 为第 i 个指标； $\bar{a}$ 为指标的平均值； n 为气象灾害指标数量。

除了从监测信息中获取的部分指标外，还需要从当地统计年鉴中获取一部分。最后将影响程度超过 80% 的指标作为最终选出的气象灾害影响指标，结果如表 1 所示^[8]。

上述选出的指标均能够在一定程度上对气象灾害进行描述，是灾害影响程度的直观体现。

1.2 气象灾害影响指标预处理

预处理的目的是提高气象灾害指标量，让其形式符合后续谱聚类算法的运算要求^[9]。

表 1 气象灾害影响指标

指标	影响程度	来源
光照强度	93.26	站点监测
气候类型	87.12	统计年鉴
植被覆盖度	85.25	统计年鉴
地理位置	81.27	统计年鉴
降水量	94.74	站点监测
地形地势	80.04	统计年鉴
日平均气温	92.14	站点监测
风力	84.12	站点监测
空气相对湿度	90.57	站点监测
受灾面积	92.33	统计年鉴
直接经济损失	94.77	统计年鉴

1) 指标标准化。

指标标准化的目标是让气象灾害指标数值映射到 0~1 之间，消除量纲^[10]。标准化公式如下：

$$c_i = \frac{a_i - \bar{a}}{b_i} \quad (2)$$

$$\bar{a} = \frac{\sum_{i=1}^n a_i}{n} \quad (3)$$

$$b_i = \sqrt{\frac{\sum_{i=1}^n (a_i - \bar{a})^2}{n}} \quad (4)$$

式中： c_i 为指标标准化结果； b_i 为第 i 个指标的方差。

2) 空位补充。

监测站在采集气象信息时，有时会存在采集不到信息，导致出现数据空缺的现象。为不影响气象灾害分布特征的提取准确性，需要进行空位补充，选取的方法为神经网络算法，通过对完整样本的训练得到空位补充模型，然后利用训练好的模型，计算得出缺失数据，完成空位补充^[11]。

3) 指标数据去噪。

在测得的指标数据中，会存在噪声的情况。噪声数据属于异常数据中的一种，在这里依据拉依达准则来进行寻找噪声，并进行噪声数据去除^[12]。具体过程如下：

步骤 1：假设重复监测指标数据 m 次。

步骤 2：计算指标数据的标准偏差。

$$S = \sqrt{\frac{\sum_{j=1}^m (a_j - \bar{a})^2}{m-1}} \quad (5)$$

式中： S 为标准偏差； a_j 为第 j 次观测指标数据。

步骤 3：计算 a_j 与平均值之间的绝对值，计算公式如下：

$$D = |a_j - \bar{a}| \quad (6)$$

步骤 4: 判断 D 是否大于等于 $3S$, 若大于等于, 认为 a_j 为异常值; 否则, a_j 为正常值。

经过标准化、空位补充以及去噪处理后得到的气象灾害指标数据基本满足谱聚类算法运算需求^[13]。

1.3 气象灾害分布特征提取模型

气象灾害分布特征提取是研究的最后一步, 主要工作是通过指标数据确定气象灾害风险程度, 完成气象灾害时空分布特征分析^[14]。在最后这个研究章节中, 气象灾害风险程度判定是关键, 所选择的方法为谱聚类算法^[15]。由此构建气象灾害分布特征提取模型, 如图 1 所示。

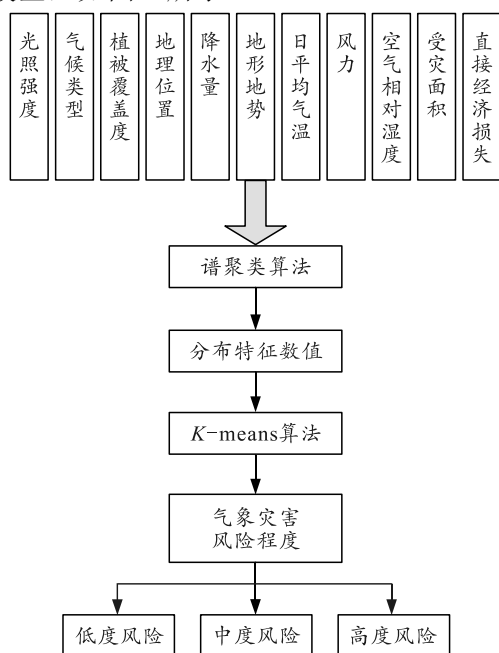


图 1 气象灾害分布特征提取模型

气象灾害分布特征提取模型应用流程如下:

- 1) 输入预处理好的气象灾害指标数据。
- 2) 将气象灾害指标数据定义为无向图。

$$Q=(R, V)。(7)$$

式中: $R=\{r_1, r_2, \dots, r_n\}$, R 为无向图顶点集合, 每个顶点就是每个指标数据; V 为顶点之间边的集合。

3) 构建 $Q=(R, V)$ 中顶点之间的相似度矩阵 L 。相似度计算公式如下:

$$L_{ij} = \begin{cases} \exp(-I_{(r_i, r_j)}^2 / 2q^2), & i \neq j \\ 0, & \text{否则} \end{cases}。(8)$$

式中: L_{ij} 为顶点 (指标数据) r_i 与 r_j 之间的相似度; $I_{(r_i, r_j)}$ 为 r_i 与 r_j 之间的欧式距离; q 为尺度参数。

- 4) 构造度量矩阵 T 。

$$T = \text{diag}(U_{11}, U_{22}, \dots, U_{ii}, \dots, U_{nn})。(9)$$

式中:

$$U_{ii} = \sum_{i,j=1}^n L_{ij}。(10)$$

- 5) 构造规范化拉普拉斯矩阵, 构造公式如下:

$$Z = T^{-0.5} \cdot L \cdot T^{-0.5}。(11)$$

式中 Z 为规范化拉普拉斯矩阵。

- 6) 提取 Z 的前 k 个特征值与对应的特征向量, 分别记为 $\{z_1, z_2, \dots, z_k\}$ 、 $\{\bar{o}_1, \bar{o}_2, \dots, \bar{o}_k\}$ 。

- 7) 由 $\{\bar{o}_1, \bar{o}_2, \dots, \bar{o}_k\}$ 构成待聚类的分布特征数值矩阵, 记为 $O = \{\bar{o}_1, \bar{o}_2, \dots, \bar{o}_k\}$ 。

- 8) 将 $O = \{\bar{o}_1, \bar{o}_2, \dots, \bar{o}_k\}$ 中行向量转变为单位向量, 得到向量空间 J 。

- 9) 将 J 的每行看作是空间中的一个点。

- 10) 执行 K -means 算法, 对 J 进行聚类, 得出气象灾害风险程度类别。具体过程如下:

步骤 1: 从 J 中随机选择 N 个向量作为初始聚类中心, 记为 w_i , $i=1, 2, \dots, N$ 。

步骤 2: 计算 J 中其余向量与这 N 个向量之间的距离, 记为 χ_{ij} 。

步骤 3: 根据 χ_{ij} 将 J 中其余向量归类给距离最近的 w_i 。

步骤 4: 计算归类后, 每类的均值, 记为 $\bar{g}_i$ 。

步骤 5: 将 $\bar{g}_i$ 作为新的聚类中心。

步骤 6: 判断聚类中心是否发生变动, 若不发生变动, 则算法结束, 完成气象灾害风险程度判定; 否则, 回到步骤 2, 按照上述过程重新进行运算^[16]。

经过上述处理后, 得出了每个监测站点所在地区的灾害风险程度, 借助 ArcGIS 将其插值到研究区地图上, 得出气象灾害分布特征, 完成基于谱聚类的畜牧业气象灾害分布特征提取模型构建^[17]。

2 实例分析

2.1 实例概况

为测试所构建模型的应用效果, 以某研究区为例, 对其 2019-2021 年 3 年间的气象灾害分布特征进行提取。研究区如图 2 所示。可看出, 该研究区内共有 10 个气象监测站点, 每个站点负责采集监测范围内的气象数据, 采集频率为每间隔 2 h 采集 1 次, 每天采集 12 次。该地区畜牧业发达, 但受到其地理位置影响, 经常发生气象灾害, 给当地畜牧业发展带来了很大的影响。为此, 进行气象灾害预警对该地畜牧业发展至关重要。

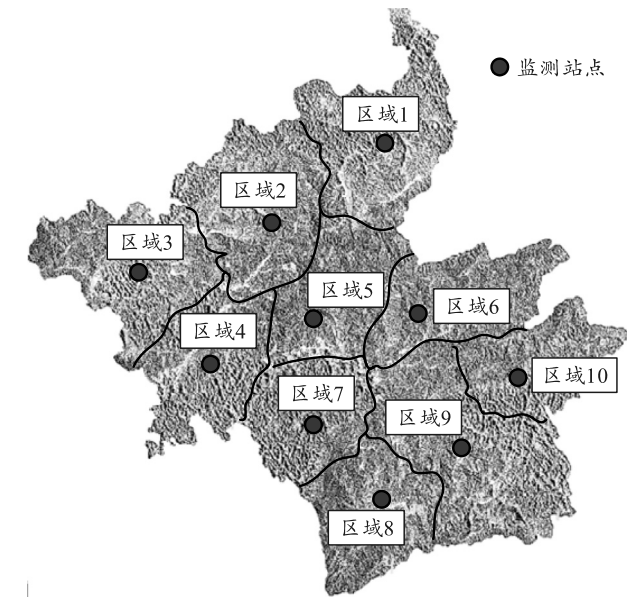


图 2 研究区

2.2 气象灾害指标数据预处理结果

针对 1.2 节研究，对气象灾害指标数据进行预处理，以其中部分样本为例，处理结果如表 2 所示。

表 2 气象灾害指标数据预处理结果示例

指标	样本 1	样本 2	样本 3
光照强度	0.552 2	0.174 5	0.212 4
气候类型	0.452 2	0.245 2	0.362 2
植被覆盖度	0.654 5	0.524 2	0.452 2
地理位置	0.336 2	0.433 3	0.477 5
降水量	0.422 2	0.458 0	0.632 1
地形地势	0.389 5	0.325 2	0.476 2
日平均气温	0.452 2	0.456 3	0.623 3
风力	0.752 2	0.632 1	0.841 2
空气相对湿度	0.355 2	0.477 1	0.621 2
受灾面积	0.245 2	0.955 2	0.552 3
直接经济损失	0.987 8	0.552 2	0.358 8

2.3 气象灾害分布特征

基于 1.3 节研究，利用谱聚类算法计算气象灾害特征值，并按照下述标准，通过 *K* 均值聚类进行气象灾害风险程度判定。

灾害特征值在 0~4 之间，气象灾害风险程度达到低度风险。

灾害特征值在 4~7 之间，气象灾害风险程度达到中度风险。

灾害特征值在 7~10 之间，气象灾害风险程度达到高度风险。

气象灾害分布特征分为 2 种，从时间角度对比，得出研究区气象灾害时间分布特征；从空间角度对比，得出研究区气象灾害时间分布特征。

1) 时间分布特征。
气象灾害分布时间特征提取结果如表 3 所示。

表 3 气象灾害时间分布特征表

区域	2019 年		2020 年		2021 年	
	灾害特征值	风险程度	灾害特征值	风险程度	灾害特征值	风险程度
1	3.532 2	低度	3.875 5	低度	5.844 2	中度
2	2.212 2	低度	4.241 4	中度	6.241 7	中度
3	3.247 1	低度	3.544 1	低度	3.475 6	低度
4	2.458 7	低度	3.647 4	低度	3.344 7	低度
5	3.466 2	低度	5.559 8	低度	7.847 2	高度
6	5.562 2	中度	3.456 4	低度	2.445 1	低度
7	2.628 5	低度	5.789 1	中度	7.899 5	高度
8	1.566 3	低度	1.854 7	低度	2.622 2	低度
9	4.784 6	中度	5.626 6	中度	5.844 9	中度
10	4.623 3	中度	4.124 1	中度	3.214 0	低度

从表 3 中可以看出，随着时间的推移，在 3 年时间内，研究区内各个区域的气象灾害呈现 3 种时间分布特征。具体如下：

① 区域 1、2、5、7 的灾害特征值随时间呈现上升趋势，说明这 4 个区域的气象灾害风险程度在逐年上升。

② 区域 6、10 的灾害特征值随时间呈现下降趋势，说明这 4 个区域的气象灾害风险程度在逐年下降。

③ 区域 3、4、8、9 的灾害特征值随时间的推移波动并不大，一直维持在较为稳定的水平上，其中 3、4、8 区域的风险程度一直保持在低度风险水平；9 区域的风险程度一直保持在中度风险水平。

2) 空间分布特征。
借助 ArcGIS 将特征空间插值到研究区地图上，得到空间分布特征图，如图 3 所示。

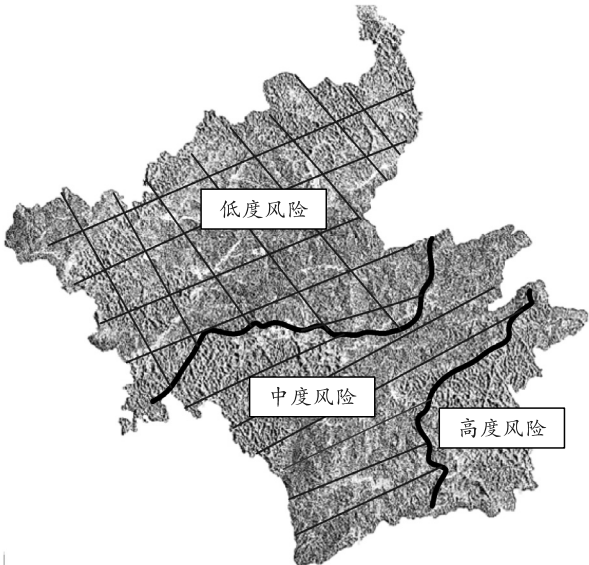


图 3 空间分布特征

从图 3 中可以看出:研究区内低度风险区域所占面积更大,且低度风险区域多集中在西北角。这是因为西北角地区地势平坦,气候、气温较为稳定,出现气象灾害的概率较小。研究区内中度风险区域所占面积排名第 2,且中度风险区域多集中在中部;最后是高度风险区域,多集中在东南地区,这一地区降水量较其他地区丰富,容易发生气象灾害。

3 结束语

笔者研究一种基于谱聚类的畜牧业气象灾害分布特征提取模型,以期实现有效的灾害预警。该研究主要是以灾害指标数据为输入,通过谱聚类得出灾害特征值以及对应的灾害风险程度,从时间和空间 2 个角度绘制图表,完成气象灾害分布特征分析。最后以某研究区为例,验证了所研究模型的应用效果,针对该区域近 3 年的灾害分布特征进行了具体分析。然而,该研究还存在不足之处。例如继续收集积累灾情数据,完善指标数据;进一步完善区划方法,制定更加详细的气象灾害规划。

参考文献:

- [1] 罗芳琼,陆虹,秦川,等. 基于 LSTM NN 特征提取的低温气象灾害预报方法研究[J]. 数学的实践与认识, 2021, 51(10): 169-182.
- [2] 赵娅君,郑粉莉,姚亚庆,等. 1978-2018 年我国农业气象灾害时空变化特征[J]. 自然灾害学报, 2022, 31(1): 198-207.
- [3] 陈朝亮,彭树宏,钱静,等. 基于 AHP-Logistic 熵权模型的西南浅丘区地质灾害分布特征研究—以内江市为例[J]. 长江科学院院报, 2020, 37(2): 55-61.
- [4] 刘琰琰,李柯薇,王鑫. 旺苍县农业气候资源及气象灾害时空特征分析[J]. 江苏农业科学, 2021, 49(4): 170-178.
- [5] 王琼,姚树然,朱锐,等. 金丝小枣生长期主要气象灾害变化特征及影响权重检验分析[J]. 干旱地区农业研究, 2021, 39(3): 202-208.
- [6] 杨辰,潘顺,严岩. 基于自然语言识别的上海市报警灾情数据识别及其气象灾害特征分析研究[J]. 自然灾害学报, 2021, 30(3): 142-150.
- [7] 于文颖,纪瑞鹏,李卓群,等. 辽宁省玉米复合农业气象灾害判识及特征[J]. 应用生态学报, 2021, 32(1): 241-251.
- [8] 惠航. 持续强降雨天气下地质危险区灾害特征提取方法[J]. 科学技术与工程, 2020, 20(14): 5667-5672.
- [9] 张辰光,邢伟,牛小燕,等. 明清时期河南省干旱灾害时空变化特征[J]. 信阳师范学院学报(自然科学版), 2021, 34(1): 88-92.
- [10] 吴炫柯,黄维,姚裕群,等. 基于 SURFER 技术的广西木薯寒冻害空间分布研究[J]. 江苏农业科学, 2020, 48(9): 103-107.
- [11] 任义方,张旭晖,陈焕根,等. 江苏河蟹高温热害时空变化特征及影响[J]. 生态学杂志, 2020, 39(11): 3704-3714.
- [12] 夏正兵. 黄河流域极端气候下降雨侵蚀力时空特征研究[J]. 水电能源科学, 2021, 39(7): 16-19.
- [13] 凌敏华,韩洪宝,吕翠美. 近 59a 河南省降水重心分布特征及其与站网空间分布关系[J]. 水利水电技术(中英文), 2022, 53(1): 26-34.
- [14] 谢晓平,刘光生. 近 60 年来福建省降雨时空分布特征[J]. 水电能源科学, 2020, 38(8): 5-8.
- [15] 韩文静,姜会飞,霍治国,等. 中国北方苹果连阴雨灾害时空分布特征[J]. 中国农业资源与区划, 2022, 43(6): 81-91.
- [16] 谢娟英,丁丽娟,王明钊. 基于谱聚类的无监督特征选择算法[J]. 软件学报, 2020, 31(4): 1009-1024.
- [17] 陶新民,王若彤,常瑞,等. 基于低密度分割密度敏感距离的谱聚类算法[J]. 自动化学报, 2020, 46(7): 1479-1495.