

环境试验与观测

大气环境分类的 SVM 联合决策方法

王竟成¹, 张伦武¹, 杨小奎¹, 胡学步², 周俊炎¹, 李泽华¹, 吴帅¹(1. 西南技术工程研究所 弹药贮存环境效应重点实验室, 重庆 400039;
2. 重庆理工大学 化学化工学院, 重庆 400054)

摘要: **目的** 实现城市大气环境的精准快速归类预测。**方法** 基于支持向量机 (SVM) 构建多分类问题的联合决策算法, 将大量城市环境因素数据的主成分聚类结果作为输入, 通过机器学习训练, 组建大气环境的 SVM 联合决策模型。**结果** 该模型根据大气环境因素将数据集 91 个城市划分为 9 类, 其中河内与海防环境相似度最高, 巴东与格尔木差异最大。9 个 SVM 二分类器组建的联合决策模型通过逐点预测在主成分数据空间形成了分区预测云图。**结论** SVM 联合决策模型可实现城市环境的快速分类辨识, 分类预测结果精度高于 95%。

关键词: 环境分类; 支持向量机; 层次聚类; 主成分分析; 联合决策; 机器学习

中图分类号: P412 文献标识码: A 文章编号: 1672-9242(2022)08-0148-07

DOI: 10.7643/issn.1672-9242.2022.08.018

SVM Joint-decision Method for Atmospheric Environment Classification

WANG Jing-cheng¹, ZHANG Lun-wu¹, YANG Xiao-kui¹, HU Xue-bu², ZHOU Jun-yan¹, LI Ze-hua¹, WU Shuai¹

(1. CSGC Key Laboratory of Ammunition Storage Environmental Effects, Southwest Technology and Engineering Research Institute, Chongqing 400039, China; 2. College of Chemistry and Chemical Engineering, Chongqing University of Technology, Chongqing 400054, China)

ABSTRACT: The paper aims to conduct a quick accurate prediction for atmospheric environment classification of different cities. SVM is used to construct a joint-decision algorithm for multi-classification problems, and the principal component clustering results of a large number of urban environmental factor data are input. Through machine learning training, the SVM joint-decision model of atmospheric environment is constructed. In the 91 cities, Hanoi and Haiphong are the most similar couple, while Padang and Golmud turn out to be the most different cities. The joint-decision model formed by 9 SVM binary classifiers forms a partitioned prediction cloud image in the principal component data space by point prediction. Results show that the prediction accuracy is higher than 95%, therefore atmospheric environment of different types can be recognized swiftly by the established model.

KEY WORDS: environment classification; SVM; hierarchical clustering; PCA; joint-decision method; machine learning

收稿日期: 2021-10-23; 修订日期: 2021-12-02

Received: 2021-10-23; Revised: 2021-12-02

基金项目: 中国博士后基金项目 (231676)

Fund: China Post Doctoral Fund Project (231676)

作者简介: 王竟成 (1991—), 男, 博士, 高级工程师, 主要研究方向为环境试验与数据挖掘分析。

Biography: WANG Jing-cheng (1991-), Male, Doctor, Senior engineer, Research focus: environmental tests and data mining.

引文格式: 王竟成, 张伦武, 杨小奎, 等. 大气环境分类的 SVM 联合决策方法[J]. 装备环境工程, 2022, 19(8): 148-154.

WANG Jing-cheng, ZHANG Lun-wu, YANG Xiao-kui, et al. SVM Joint-decision Method for Atmospheric Environment Classification[J]. Equipment Environmental Engineering, 2022, 19(8): 148-154.

迈入数据时代, 时刻都有千亿量级的数据产生, 高效的数据分析处理手段一直都是研究热点。气象环境是典型的海量数据生产制造领域, 遍布全球的气象监测站不断收集着温度、湿度、降水等数据。这些宝贵的数据资源在经济、军事、社会等各个领域产生着巨大效益。运用各类挖掘分析算法研究环境数据资源的内在规律, 提炼隐性知识, 能显著提升我们的认知, 有效促进数据增值。

气候分类最初源于农业生产, 通常对应于植被分布, 即每种气候类型由一种植被区域或生态区域支配^[1]。气候分类的方法很多, 其中科蓬气候分类及其衍生方法^[2]应用最多。温度和降水是植被分布的主导性因素, 因而是气候分类最基本的变量^[3]。材料/产品的使用、服役与植被的分布类似, 也具有一定的区域局限性, 并且除温度、降水外, 应考虑更多的环境因素, 如湿度、太阳辐射等。太阳辐射会显著加速高分子材料的老化, 湿空气极易引起电子元器件的失效, 此外还有温度-湿度-太阳辐射之间的耦合作用引发的性能退化与功能失效。基于农学的传统气候分类方法, 不能满足装备产品的多环境因素的分类辨识需求。因此, 构建一个综合的环境分类分析决策方法指导装备设计与使用维护具有重要意义。

诸多学者对大气环境的分类辨识开展了研究。如张伦武等^[4]使用模糊聚类分析了腐蚀大气环境。吴超等^[5]采用有序样本聚类开展了腐蚀大气分类的细化研究。唐其环^[6]使用灰色聚类分析了化学成分对大气腐蚀的影响。前期研究中^[7], 从产品使用与装备服役的角度, 提出了主成分聚类分析法, 用于中国城市的环境分类, 但并未考虑新增城市的快速分类辨识问题。本文进一步进行了深化拓展, 以 91 个亚非欧城市的环境数据作为输入, 在主成分聚类分析的基础上, 采用多个支持向量机分类器, 组建大气环境分类的联合决策模型, 用于城市环境的分类辨识, 指导产品设计与使用维护。

1 数据与方法

1.1 技术路线

本研究构建了一种联合决策模型用于城市大气环境的快速分类辨识, 技术路线流程如图 1 所示。首先收集、整理能代表地区环境类型的环境因素, 具体包含温度、相对湿度、降水量和太阳辐射累积量等, 组装成分析数据矩阵。将各数据列标准化后进行主成分分析, 计算得到特征值与特征向量, 利用累积贡献率确定主要成分, 采用主成分的欧几里得距离表征各城市之间的相似性, 以此为依据进行层次聚类。将聚类结果作为训练测试数据, 利用联合支持向量机 (Support Vector Machine, SVM) 分类器进行得分评估, 从而构建联合决策模型。采用训练好的联合决策

模型, 不仅能对新增城市的大气环境类型进行判别, 也能对大气环境的主成分数据空间进行区域划分, 形成分区云图。

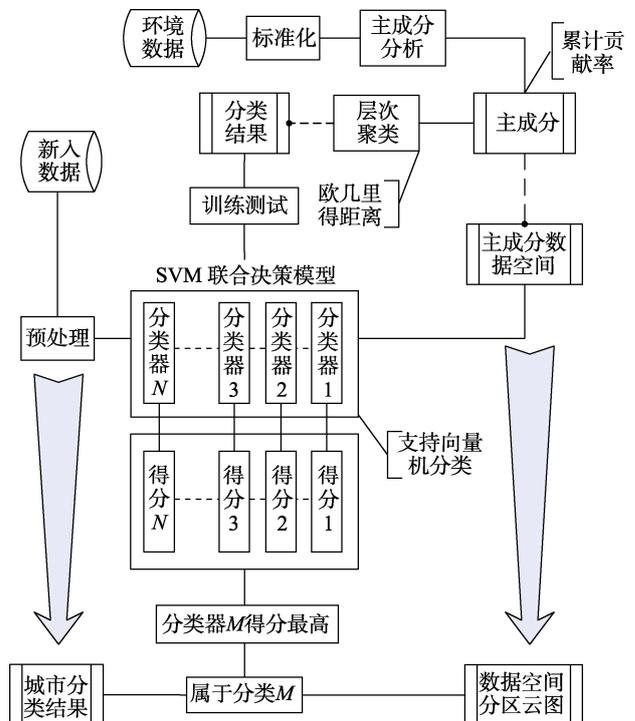


图 1 技术路线流程

Fig.1 Technical route flow diagram

1.2 数据

为开展国内外城市大气环境的分类预测研究, 从中国气象数据中心网站下载了 91 个亚非欧城市 (包含 38 个中国城市) 近 10 a (2010—2019 年) 的年度数据。数据集包含 5 个有效环境变量, 包括纬度、年平均温度、年均相对湿度、年均降水量、年累积辐射量。求取各环境因素 10 a 间年度数据的平均值, 组合成分析矩阵, 以此表征这些城市的大气环境信息, 部分城市环境数据见表 1。

表 1 部分国内外城市环境数据
Tab.1 Environmental data of 8 cities

城市	纬度/ (°)	平均温度/ °C	相对湿度/ %	降水量/ mm	辐射累积量/ (MJ·m ⁻²)
/雅典	38	18.5	69.7	514.7	5 958.6
开罗	30.1	22.1	46.9	51.1	7 389.1
孟买	19.0	27.1	71.8	2 022.1	7 373.0
格尔木	36.4	6.6	31.6	55.4	8 007.7
巴东	-1.0	27.2	84.5	3 668.3	6 156.1
北京	39.8	13.4	51.8	578.8	6 282.0
河内	21.0	23.1	80.3	1 569.5	5 100.5
海防	20.9	23.8	81.6	1 649.8	5 159.6

表 1 中, 各列环境变量之间并非完全独立, 存在一定内在联系。例如, 平均温度明显受纬度影响, 相对湿度和降水显著相关。将表中原始数据 (X_i) 以列为对象标准化后 (X'_i), 采用式 (1) 计算相关系数矩阵 $[R]$ 。

$$[R_{ij}]_{p \times p} = \frac{\text{cov}(X_i, X_j)}{\text{var}(X_i) \cdot \text{var}(X_j)} = \text{cov}(X'_i, X'_j) \quad (1)$$

式中: var 表示方差, cov 表示协方差

由表 2 的相关系数矩阵可见, 平均温度、降水量与纬度呈负相关, 相关系数分别为 -0.821、-0.793。降水量与湿度相关系数为 0.803, 显著正相关。辐射累积量与纬度、平均温度的相关性很低。

表 2 相关系数矩阵

Tab.2 Correlation coefficient matrix

	纬度	平均温度	相对湿度	降水量	辐射累积量
纬度	1.000	-0.821	-0.586	-0.793	-0.128
平均温度	-0.821	1.000	0.380	0.593	0.117
相对湿度	-0.586	0.380	1.000	0.803	-0.484
降水量	-0.793	0.593	0.803	1.000	-0.239
辐射累积量	-0.128	0.117	-0.484	-0.239	1.000

1.3 分析方法与模型

多变量数组中, 一些变量之间存在随动关系, 可能是衡量同一系统行为的因素。在气象数据系统中, 诸多变量仅有少数的驱动型因素, 因此可以采用少数的组合变量来代替原始变量组, 即进行数据降维, 以此简化数据分析矩阵。主成分分析法 (Principal Component Analysis, PCA) 作为一种广泛应用^[8-12]的高效算法, 能有效地从多变量数据中剔除冗余成分, 筛选少数足以解释结果的独立成分。本研究将表征城市环境特征的数据矩阵通过 PCA 提取主成分, 作为层次聚类的输入。

聚类是数据分析中最基本、最普遍的问题。观测对象数据通常自然成团聚类, 同一类的对象特征相似, 而不同类的对象特征相异。本研究采用层次聚类, 通过构建聚类树将不同对象的数据进行聚集。聚类树并非如同 K-Means 方法中的单个聚集体, 而是一个多层级的层次结构, 上个层级的聚集体在下个层级会继续合并在一起, 各对象之间的相互关联较清晰。在具体应用时, 可以根据需求选择最合适的聚类层级或规模。研究中首先计算各城市主成分在数据空间中的距离, 评估其相似性。根据相似性, 将关系相近的对象归类到二元聚集体中, 新的聚集体又进一步归类到更大的聚集体, 直至聚类树形成。以聚类树为依据, 选择合理的层级, 得到城市大气环境的聚类结果。

将上述主成分聚类分析的结果划分成训练集与测试集, 采用支持向量机 (SVM) 算法进行分类学习训练与预测。支持向量机的主要思想是将输入空间的

非线性问题, 通过核函数映射到高维特征空间, 并在此空间搜寻使类间间隔最大的超平面, 将数据样本划分开。目前常用的核函数主要有多项式、高斯径向基函数、Sigmoid、傅里叶等 4 种核函数。基于结构风险最小化准则, SVM 网络拓扑结构由支持向量决定, 是介于简单算法和神经网络之间最好的算法。它仅有 3 个自由参数, 问题复杂度不取决于特征的维数, 只通过几个支撑向量就能确定超平面, 忽略细枝末节, 可以处理复杂的非线性问题。同时, 通过引入松弛变量, SVM 能够解决类间的重叠问题, 并提高了泛化能力。SVM 在处理小样本、非线性时具有明显优势, 可规避其他机器学习算法中易于出现的局部极小以及过拟合现象, 已被广泛应用于趋势预测^[13-14]与故障诊断等领域^[15], 如空气质量预测^[16-20]。

根据城市的分类数确定分类器个数, 由各个分类器的得分评判城市的所属分类, 并采用测试数据评估训练后模型的有效性。串联起 PCA、层次聚类与 SVM 分类器, 组建大气环境分类的 SVM 联合决策模型, 模型部署后可用于城市大气环境分类的快速判别与主成分数据空间的区域划分。

2 结果与讨论

2.1 主成分聚类分析

由相关系数矩阵 (见表 2), 构建特征矩阵 $|\lambda I - R| = 0$, 计算得到特征值及其对应的特征向量 (见表 3)。特征值表征成分之间的方差, 采用降序排列, 每个特征向量即为一个主成分的参数矢量。由表 3 可见, 主成分 1 (PC_1) 方差的贡献率为 60.6%, 主成分 2 (PC_2) 的贡献率为 27.3%, 主成分 3 (PC_3) 的贡献率为 7.5%。仅采用 PC_1 与 PC_2 时, 方差的累积贡献率为 87.9%; 而采用前 3 个主成分时, 累积贡献率可高达 95.4%。

表 3 特征值及其对应特征向量

Tab.3 Eigenvalues and corresponding eigenvector

主成分	特征值	贡献率/%	特征向量 B_i
1	3.03	60.6	-0.521, 0.449, 0.475, 0.536, -0.117
2	1.37	27.3	-0.299, 0.373, -0.380, -0.094, 0.786
3	0.37	7.5	0.053, 0.695, -0.366, -0.321, -0.526
4	0.14	2.9	0.052, 0.186, 0.689, -0.675, 0.183
5	0.09	1.7	0.796, 0.376, 0.148, 0.381, 0.241

将特征向量 (B_i) 与标准化后的变量值 (X'_i) 点乘计算主成分得分 PC_i 。本研究中, 选择前 3 个主成分来表征城市的主要环境特征, 包含原始变量 95.4% 的信息。以主成分得分作为坐标值 (PC_1, PC_2, PC_3), 在图 2 的数据空间中把每个城市标注出来, 根据距离远近粗略地展示了不同城市之间的相似性。

表 4 城市大气环境聚类结果

Tab.4 Atmospheric environment classification of 91 cities

类别	城市
类型 1	漠河
类型 2	吉尔吉特 安卡拉 伊兹密尔 安塔基亚 北京 天津 太原 沈阳 长春 哈尔滨 济南 昌都 西宁 阿勒泰
类型 3	拉萨
类型 4	二连浩特 敦煌 格尔木 银川 喀什 哈密 吐鲁番
类型 5	河内 海防 荣市 杭州 福州 南昌 武汉 长沙 广州 桂林 南宁 海口 万宁 重庆 成都 贵阳
类型 6	巴黎 伊斯坦布尔 特拉布宗 雅典 上海 南京 合肥 郑州 昆明 西安
类型 7	曼谷 清迈 孟买 班加罗尔 海德拉巴 博帕尔 加尔各答 巴特那 胡志明市 芽庄 阿克拉 合艾 素叻 雅加达 泗水 万隆 棉兰 巨港 巴东
类型 8	果阿 峴港 吉隆坡 乔治市 马六甲 亚庇 古晋市 斯里巴加湾市 科伦坡 康提'
类型 9	新德里 昌迪加尔 艾哈迈达巴德 坎普尔 阿布扎比 迪拜 伊斯兰堡 卡拉奇 奎达 图尔巴特 耶路撒冷 开罗 阿斯旺

利用不同城市的主成分得分进行层次聚类, 通过计算欧几里得距离 $d=\sqrt{PC_1^2+PC_2^2+PC_3^2}$, 衡量城市之间的相似性, 从而构建层次聚类树, 如图 3 所示。从聚类树可见, 当前 91 个城市大气环境的最优聚类数目为 9, 由差异值等于 2.5 的虚线从层次聚类树中划分出来。虚线之下, 同一簇团内城市之间的相异性低于 2.5, 代表一种环境类型, 用同种颜色绘制。91 个城市中, 海防与河内的差异值最低, 仅为 0.135, 从大气环境的角度来说它们犹如孪生城市; 格尔木与巴东之间差异值最高, 为 6.701。城市聚类结果见表 4, 类型 8 为赤道附近的热带雨林气候, 类型 7 为高温多雨的热带气候, 类型 5 为亚热带湿润气候, 类型 9 为亚热带半干旱气候, 类型 6 为温带湿润气候, 类型 4 为温带干旱气候, 类型 2 为寒温带半湿润气候。严寒地区漠河 (类型 1)、高海拔城市拉萨 (类型 3) 与其他城市的环境特征区别较大, 自成一类。

2.2 SVM 联合决策分类

基于支持向量机建立 9 个并行的二分类器, 每个

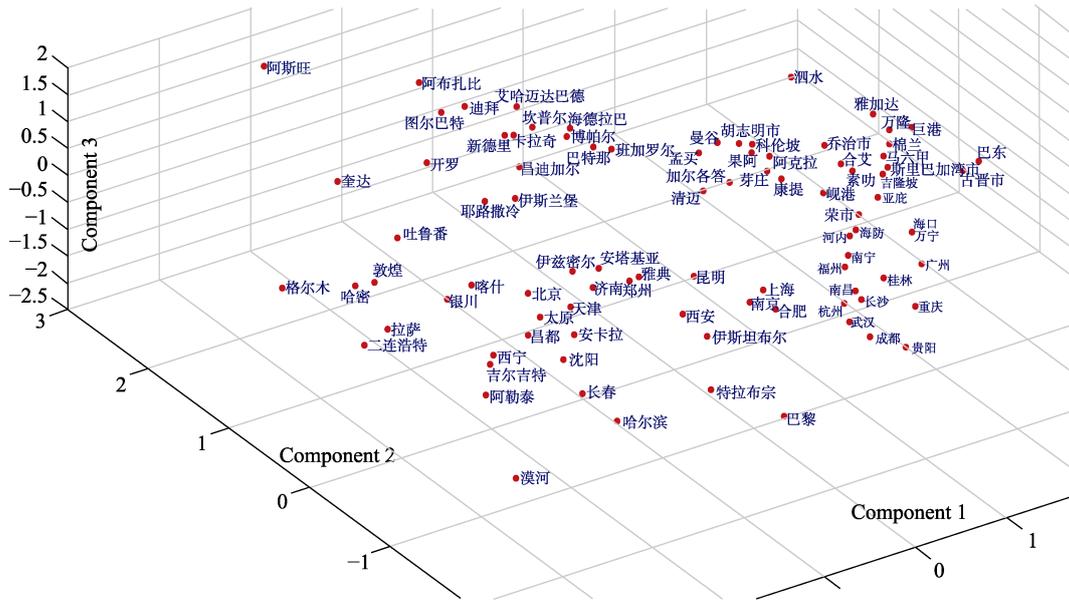


图 2 主成分数据空间中的城市分布

Fig.2 Distribution of 91 cities in principal component data space

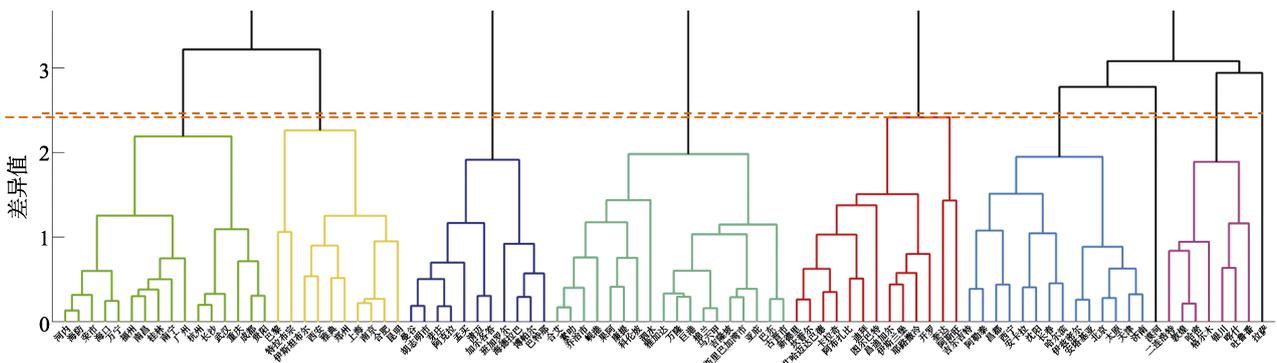


图 3 91 个城市的层次聚类树

Fig.3 Hierarchical clustering tree of 91 cities

分类器评估对应编号类别的可信度,使用分区聚类结果对所有分类器进行训练。利用训练好的联合分类器分别计算城市大气环境属于每个分区的可信度得分,指定可信度得分最高的分类为当前城市的归属分区,并根据得分判断其可信水平。对于当前应用场景,规定:可信度>1,可信水平极高;1>可信度>0.4,可信水平高;0.4>可信度>0,可信水平中;0>可信度>-0.4,可信水平低;-0.4>可信度,不可信。

91个城市中,抽取10%(9个)的城市作为测试集,其余城市作为训练集训练SVM联合决策模型。由于类型1与类型3仅有1个城市(见表4),为保证训练集对象种类的全覆盖,应避免将拉萨与漠河选入测试集。从测试结果(见表5)来看,本文构建的SVM联合决策模型对当前数据集的预测准确率为100%。从可信度得分来看,9个测试数据中4个城市的可信度>1,可信度极高;3个城市可信度∈[0.4, 1],可信度高;2个城市可信度∈[0, 0.4]。其中,可信水平高以上的占比为77.8%。可见,大气环境分类的SVM联合决策模型具有较高的可靠性。

表5 测试集分类预测结果
Tab.5 Prediction results of test data

城市	原始/预测分类	正误	可信度	可信水平
雅加达	8/8	○	0.891	高
阿布扎比	9/9	○	1.241	极高
芽庄	7/7	○	0.227	中
乔治市	8/8	○	1.314	极高
伊斯兰堡	9/9	○	0.231	中
天津	2/2	○	0.509	高
合肥	6/6	○	0.497	高
南宁	5/5	○	1.171	极高
哈密	4/4	○	1.171	极高

将91个城市的数据全部用于训练,生成样本点更全面的联合决策模型。统计结果(表6)显示,其预测准确度97.8%,预测博帕尔与巴纳特时失效,这2个城市聚类结果为类型7,预测结果为类型9,可信水平为中。在后文的主成分数据空间中,这2个城市正好处于分区7与分区9的交界面,说明博帕尔与巴纳特处于类型7向类型9过渡的区域。在图3的层次聚类树中,若聚类的相异值设定为2,那么博帕尔、巴纳特、海德拉巴与班加罗尔将从蓝色分支(类型7)中独立出来,成为一个新的细分类型。预测结

表6 分类决策模型的有效性
Tab.6 Validity of present prediction model

可信度	低	中	高	极高	失效
城市数	3	12	21	53	2
占比/%	3.3	13.2	23.1	58.2	2.2
总计/%		97.8			2.2

果中,可信度极高的数目占比为58.2%,高以上的占81.3%。由此可见,本文所构建的模型既具有一定泛化能力,没有出现学习现象,也保持了较高的准确率与可靠性。

限定边界框范围为: $\min(PC_1) \leq X \leq \max(PC_1)$, $\min(PC_2) \leq Y \leq \max(PC_2)$, $\min(PC_3) \leq Z \leq \max(PC_3)$, 遍历主成分数据空间边界框内的所有点,格点间隔取0.1。通过分类决策模型计算每个格点的可信度,将可信度得分最高的类别判定为该格点的分类。对于最高可可信度得分低于-0.4的格点,判定为不可信,不划入任何分类,数据空间中表现为空白区域。图4将各分类的区域范围可视化为不同颜色的团簇。取 $PC_3=0$, 绘制分区云图中间截面的剖面图(图4右上部),包含7个分区。在分区云图中直观可见各分区之间的相互关联:热带气候(分区7)处于数据空间的中心位置,与大多数环境类型都相接;热带雨林气候(分区8)与温带干旱气候(分区4)差异很大,处于数据空间的另一端;严寒地区(分区1)与高海拔地区(分区3)处于数据空间的边缘位置,环境典型而独立。

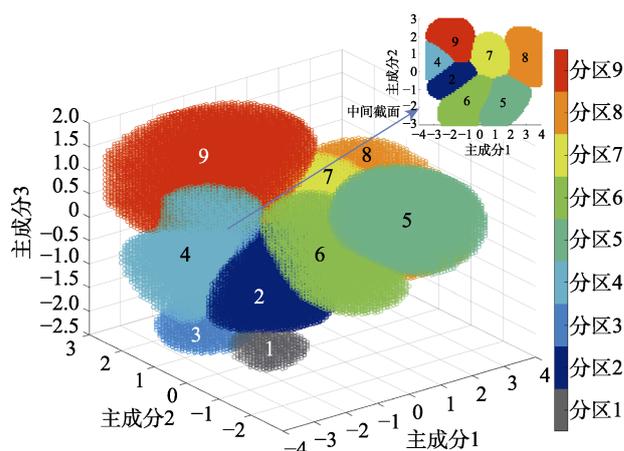


图4 主成分数据空间中的分区云图
Fig.4 Classification clusters in principal component data space

由于数据集中分区1与分区3仅有1个城市,通过联合决策模型判别出的区域较小,处于主成分数据空间左下部($PC_1 < 0, PC_3 < 0$)。数据空间右下部($PC_1 > 0, PC_3 < 0$)表征着中等温度、中等辐射、高湿度、高降水的环境,该区域为空白表明当前数据集中不含这类型的城市。数据空间右上部($PC_1 > 0, PC_3 > 0$)也是空白,这块区域代表着高温、低辐射、中等湿度、中等降水环境,而高温、低太阳辐射这种环境类型比较特殊,满足此条件的有中高纬度的活火山附近地区,不适宜人类居住。

3 结论

1)采用主成分聚类分析将亚非欧91个城市划分

为 9 类大气环境类型, 3 分量主成分包含了 5 种环境因素 95.4% 的信息, 数据集城市中河内与海防环境最类似, 巴东与格尔木环境差异最大。

2) 融合 9 个支持向量机分类器组建联合决策模型, 利用主成分聚类结果对模型进行训练, 结果表明, 该模型对大气环境类型预测的准确率高达 97.8%。

3) 应用大气环境分类的联合决策模型将主成分数据空间划分为 9 大团簇, 并可视化为分区云图, 使各分区之间的相互关联直观可见。热带气候处于数据空间的中心位置, 与大多数环境类型相接; 热带雨林气候与温带干旱气候差异很大, 处于数据空间两端; 严寒地区与高海拔地区处于数据空间的边缘位置, 环境典型而独立。

参考文献:

- [1] MICHAL B, EVA H. Climate Classification Revisited: From Köppen to Trewartha[J]. *Climate Research*, 2014, 59(1): 1-13.
- [2] PEEL M, FINLAYSON B, MCMAHON T. Updated World Map of the Köppen-Geiger Climate Classification[J]. *Hydrology and Earth System Sciences*, 2007, 11: 1633-1644.
- [3] SPAROVEK G, VAN LIER Q D J, DOURADO NETO D. Computer Assisted Koeppen Climate Classification: A Case Study for Brazil[J]. *International Journal of Climatology*, 2007, 27(2): 257-266.
- [4] 张伦武, 汪学华, 封先河, 田月娥, 王成章. 大气腐蚀环境及环境因素模糊聚类分析[J]. *中国腐蚀与防护学报*, 2003, 23(2): 112-115.
ZHANG Lun-wu, WANG Xue-hua, FENG Xian-he, et al. The Environment of Atmospheric Corrosion and Environmental Factors Fuzzy Clustering Analyze[J]. *Journal of Chinese Society for Corrosion and Protection*, 2003, 23(2): 112-115.
- [5] 吴超, 付冬梅, 李晓刚. 基于数据分析的大气腐蚀等级细化研究[J]. *装备环境工程*, 2015, 12(4): 16-21.
WU Chao, FU Dong-mei, LI Xiao-gang. Refinement Research of Atmospheric Corrosion Categories Based on Data Analysis[J]. *Equipment Environmental Engineering*, 2015, 12(4): 16-21.
- [6] 唐其环. 从灰色聚类结果看化学成分对钢大气腐蚀的影响[J]. *中国腐蚀与防护学报*, 2003, 23(4): 239-242.
TANG Qi-huan. The Effects of Chemical Composition to the Atmospheric Corrosion of Steel—gained from Grey Clustering Analysis[J]. *Journal of Chinese Society for Corrosion and Protection*, 2003, 23(4): 239-242.
- [7] WANG JINGCHENG, ZHANG LUNWU, ZHANG DINGFEI, et al. The Application of Principal Component Cluster Analysis in Environment Classification for Chinese Cities[J]. *IOP Conference Series Earth and Environmental Science*, 2020, 569(1): 012040.
- [8] 贾志军, 赵泉林, 董超芳, 等. 主成分分析法综合评价我国 13 个大气腐蚀站点的气候因素[J]. *中国腐蚀与防护学报*, 2009, 29(5): 388-393.
JIA Zhi-jun, ZHAO Quan-lin, DONG Chao-fang, et al. Evaluation of Weathering Conditions of 13 Cities in China by Principal Component Analysis[J]. *Journal of Chinese Society for Corrosion and Protection*, 2009, 29(5): 388-393.
- [9] 方红卫, 孙世群, 朱雨龙, 等. 主成分分析法在水质评价中的应用及分析[J]. *环境科学与管理*, 2009, 34(12): 152-154.
FANG Hong-wei, SUN Shi-qun, ZHU Yu-long, et al. Principal Component Analysis Apply in Water Quality Assessment[J]. *Environmental Science and Management*, 2009, 34(12): 152-154.
- [10] 陈军辉, 谢明勇, 傅博强, 等. 西洋参中无机元素的主成分分析和聚类分析[J]. *光谱学与光谱分析*, 2006, 26(7): 1326-1329.
CHEN Jun-hui, XIE Ming-yong, FU Bo-qiang, et al. Principal Component Analysis and Cluster Analysis of Inorganic Elements in Panax Quinquefolium.L[J]. *Spectroscopy and Spectral Analysis*, 2006, 26(7): 1326-1329.
- [11] 宋伟, 熊伟, 董莎莎, 等. 多元时间序列 PCA 分割及在天然气脱水装置工况识别中的应用[J]. *装备环境工程*, 2020, 17(4): 85-89.
SONG Wei, XIONG Wei, DONG Sha-sha, et al. PCA Segmentation of Multivariate Time Series and Its Application in the Condition Identification of Natural Gas Dehydration Unit[J]. *Equipment Environmental Engineering*, 2020, 17(4): 85-89.
- [12] 熊伟, 马豪, 张波, 等. PCA 和 SDG 融合的天然气管道脱水装置故障诊断[J]. *装备环境工程*, 2021, 18(9): 107-113.
XIONG Wei, MA Hao, ZHANG Bo, et al. Fault Diagnosis for Natural Gas Dehydration Device Based on the Combination of PCA and SDG[J]. *Equipment Environmental Engineering*, 2021, 18(9): 107-113.
- [13] 冉林, 熊建军, 赵照, 等. 基于机器学习的电热防除冰表面温度变化趋势预测[J]. *装备环境工程*, 2021, 18(8): 29-35.
RAN Lin, XIONG Jian-jun, ZHAO Zhao, et al. Prediction of Surface Temperature Change Trend of Electric Heating Anti-Icing and De-Icing Based on Machine Learning[J]. *Equipment Environmental Engineering*, 2021, 18(8): 29-35.
- [14] 徐安桃, 李锡栋, 周慧. 基于 SOM-SVM 组合分类器的涂层防护性能研究[J]. *装备环境工程*, 2018, 15(5): 62-66.
XU An-tao, LI Xi-dong, ZHOU Hui. Protective Performance of Coating Based on Self-Organizing Neural Network and Support Vector Machine[J]. *Equipment Environmental Engineering*, 2018, 15(5): 62-66.
- [15] 唐其瑄, 车驰东. 基于决策树支持向量机算法的船用设备故障诊断[J]. *装备环境工程*, 2021, 18(9): 72-77.
TANG Qi-xuan, CHE Chi-dong. Research on Marine Equipment Fault Diagnosis Based on Decision Tree Sup-

- port Vector Machine Algorithm[J]. *Equipment Environmental Engineering*, 2021, 18(9): 72-77.
- [16] 宋国君, 国潇丹, 杨啸, 等. 沈阳市 $PM_{2.5}$ 浓度 ARIMA-SVM 组合预测研究[J]. *中国环境科学*, 2018, 38(11): 4031-4039.
- SONG Guo-jun, GUO Xiao-dan, YANG Xiao, et al. ARIMA-SVM Combination Prediction of $PM_{2.5}$ Concentration in Shenyang[J]. *China Environmental Science*, 2018, 38(11): 4031-4039.
- [17] 王黎明, 吴香华, 赵天良, 等. 基于距离相关系数和支持向量机回归的 $PM_{2.5}$ 浓度滚动统计预报方案[J]. *环境科学学报*, 2017, 37(4): 1268-1276.
- WANG Li-ming, WU Xiang-hua, ZHAO Tian-liang, et al. A Scheme for Rolling Statistical Forecasting of $PM_{2.5}$ Concentrations Based on Distance Correlation Coefficient and Support Vector Regression[J]. *Acta Scientiae Circumstantiae*, 2017, 37(4): 1268-1276.
- [18] 李龙, 马磊, 贺建峰, 等. 基于特征向量的最小二乘支持向量机 $PM_{2.5}$ 浓度预测模型[J]. *计算机应用*, 2014, 34(8): 2212-2216.
- LI Long, MA Lei, HE Jian-feng, et al. $PM_{2.5}$ Concentration Prediction Model of Least Squares Support Vector Machine Based on Feature Vector[J]. *Journal of Computer Applications*, 2014, 34(8): 2212-2216.
- [19] 刘杰, 杨鹏, 吕文生, 等. 模糊时序与支持向量机建模相结合的 $PM_{2.5}$ 质量浓度预测[J]. *北京科技大学学报*, 2014, 36(12): 1694-1702.
- LIU Jie, YANG Peng, LÜ Wen-sheng, et al. Prediction Model of $PM_{2.5}$ Mass Concentrations Based on Fuzzy Time Series and Support Vector Machine[J]. *Journal of University of Science and Technology Beijing*, 2014, 36(12): 1694-1702.
- [20] 苏筱倩, 安俊琳, 张玉欣. 基于支持向量机回归和小波变换的 O_3 预报方法[J]. *中国环境科学*, 2019, 39(9): 3719-3726.
- SU Xiao-qian, AN Jun-lin, ZHANG Yu-xin. Support Vector Machine Regression Forecasting of O_3 Concentrations Based on Wavelet Transformation[J]. *China Environmental Science*, 2019, 39(9): 3719-3726.

责任编辑: 刘世忠