

基于 BiLSTM 改进聚类的空气质量监测点位优化^①



李 幔^{1,2}, 马元婧¹

¹(中国科学院 沈阳计算技术研究所, 沈阳 110168)

²(中国科学院大学, 北京 100049)

通信作者: 李 幔, E-mail: 1370794464@qq.com

摘 要: 近年来, 空气质量监测微子站监测逐渐成为了空气质量监测网络的重要组成部分. 随着经济的不断发展, 城市化进程的不断加快, 站点的冗余以及代表性降低的问题逐渐显现. 由于空气质量监测微子站抵抗突发环境因素能力较弱, 极易导致监测数据缺失, 不仅会大大增加数据分析的复杂性与难度, 还会导致优化布点结果的偏差. 本文针对以上问题, 提出了一种将 BiLSTM 神经网络结合聚类的点位优化方法, 在应用 BiLSTM 神经网络补全缺失数据的基础上, 应用凝聚层次聚类法对修复后的数据进行聚类. 在实现用尽可能少而准确的点位反馈空气质量水平的基础上, 大大提升聚类准确度. 最后, 本文使用沈阳市位于浑南区的 18 个空气质量监测微子站的监测数据进行实验验证. 结果表明, 相比于一般的聚类算法, 本文提出的算法性能有一定提升, 为空气质量监测点位优化提供了一种新方法.

关键词: 空气质量监测; BiLSTM 神经网络; 数据缺失; 点位优化; 聚类分析

引用格式: 李幔, 马元婧. 基于 BiLSTM 改进聚类的空气质量监测点位优化. 计算机系统应用, 2022, 31(6): 217-223. <http://www.c-s-a.org.cn/1003-3254/8521.html>

Optimization of Air Quality Monitoring Site by Improved Clustering Method Based on BiLSTM

LI Man^{1,2}, MA Yuan-Jing¹

¹(Shenyang Institute of Computing Technology, Chinese Academy of Sciences, Shenyang 110168, China)

²(University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: In recent years, mini-station monitoring of air quality (AQ) has gradually become an important part of the AQ monitoring network. With the continuous economic development and increasing acceleration of the urbanization process, monitoring stations appear redundant and not as representative as they used to be. Due to the weak ability of AQ monitoring mini-stations against sudden environmental factors, monitoring data missing can easily occur, which not only greatly increases the complexity and difficulty of data analysis but also leads to the deviation of optimization and distribution results. In view of the above problems, we propose a site optimization method that combines the BiLSTM neural network with clustering, of which the BiLSTM neural network is applied to complete the missing data, and the condensed hierarchical clustering method is employed for the clustering of restored data. On the basis of the feedback on the AQ level with as few and accurate sites as possible, the clustering accuracy is greatly improved. Finally, this study uses the monitoring data from 18 AQ monitoring mini-stations in the Hunnan District of Shenyang for verification. The results show that the performance of the proposed algorithm is improved to a certain extent compared with that of general clustering algorithms, which provides a new method for the optimization of AQ monitoring sites.

Key words: air quality monitoring; BiLSTM neural network; data missing; site optimization; clustering analysis

① 基金项目: 国家水体污染控制与治理科技重大专项 (2018ZX07601001)

收稿时间: 2021-09-07; 修改时间: 2021-09-29; 采用时间: 2021-10-14; csa 在线出版时间: 2022-05-26

随着我国经济水平的提高,交通运输业的飞速发展以及工业现代化进程的不断加快,空气质量水平作为可持续发展的重要组成部分,也越来越受到人们的关注.为及时精准地了解空气质量问题时空分布状况,提高动态反应空气质量的能力,空气质量监测应运而生.根据多年来的实验表明,环境监测在采样过程中产生的误差是总体误差的重要组成部分^[1].因此,保证采样过程中监测点位选择的准确度就成为了空气质量监测与表征空气污染程度的关键^[2].而监测点位的优化准确度依赖于空气质量监测微子站(以下简称为微子站)的监测数据的有效完整性,但在微子站监测过程中,常常由于各种突发因素导致一段时间内的数据缺失.因此,在根据微子站监测数据进行点位优化前,进行缺失数据处理成为了不可或缺的一部分.

近年来,随着空气质量监测点位优化问题研究的不断深入,已有众多发达国家与地区对于空气质量监测点位优选提出了一系列的优化方法,如相关性分析法、聚类分析法以及多目标优化等数学与统计学方法^[3].我国对于空气质量监测点位优化的研究也在逐步发展,如徐明德等人提出的凝聚层次物元法^[4]、杜增荣采用的凝聚层次聚类法^[5]、姜林等人提出的BP分析法^[6]应用于空气质量监测点位优化.这些方法各有特色,可以很好地代表空气质量的准确性与代表性.与此同时,随着对缺失数据处理问题的关注度不断提高,大多被提出的数据填充方案大概分为以下两种填充思想,第一种是对缺失数据邻近数据进行特征分析,利用均值,众数和聚类的方法表现数据特征进行填补的算法.如李董等人提出的基于SMOTE和KNN的数据缺失填充算法^[7]、袁兆祥等人提出的基于DBSCAN聚类的数据补充算法^[8]、杨挺等人提出的基于FSOM神经网络的数据补充方案^[9],另一种缺失数据的填补方式是利用神经网络对数据整体特征进行分析与补充,例如使用RNN来对缺失的数据进行填补操作.如柯昊等人提出的BP神经网络补差法^[10]、宋维等人提出的基于LSTM的活立木茎干水分缺失数据填补方法^[11],对于数据缺失问题都有很好的表现.因此,本文首先选取在空气质量、水质、地质预测领域应用广泛,原理简单,计算量小的凝聚层次聚类法进行空气质量监测点位优化.鉴于凝聚层次聚类法易受到初始聚类中心以及缺失数据的影响,单独使用凝聚层次聚类法的聚类效果不理想.为了能够在空气质量监测点位监测数据缺失的情况下,

拥有更好的聚类效果,本文首先选取适用于具有时序特征缺失数据补充且能够反映空气质量监测数据双向特征的BiLSTM神经网络对缺失数据进行填补,然后应用凝聚层次聚类法对修复后的数据进行聚类分析,实现对空气质量监测点位的优化,获取更加真实、精准的数据,用以分析空气污染成因.同时,为环境质量监测有关部门制定点位优化标准提供有利的技术支撑.

1 基于BiLSTM改进聚类的点位优化模型

1.1 模型整体框架

针对空气质量监测点位优化的最终目的,构建的基于BiLSTM改进聚类的空气质量监测点位优化模型,如图1所示.首先,利用BiLSTM数据补充模型进行数据补充解决监测数据缺失问题,对于BiLSTM中的正向LSTM网络和逆向LSTM网络具备相同的结构,LSTM_L部分是通过学习缺失数据之前时序的空气质量数据的数据特征规律来产生缺失数据的候选补充结果;LSTM_R是通过学习缺失数据之后时序的空气质量数据进行同样的操作,产生另一候选补充结果.BiLSTM神经网络通过使用缺失监测数据的前后 m 组数据对当前缺失的 n 组数据进行补充,公式如下:

$$[X_k \cdots X_{k+n-1}] = F \left(\begin{array}{c} X_{k-m} \cdots X_{k-1} \\ X_{k+n+m} \cdots X_{k+m+1} \end{array} \right)$$

其中, F 为基于BiLSTM的数据补充模型, $[X_k \cdots X_{k+n-1}]$ 表示缺失的 n 组数据.

最后使用一个Softmax layer来综合正向LSTM和逆向LSTM的候选结果并产生最终的缺失数据补充结果.得到最终的输出结果传递给层次聚类部分,输出聚类结果.

1.2 BiLSTM模块

LSTM神经网络是一种循环神经网络(RNN)的特殊变体.在标准的RNN中,重复部分的模块表现为只有单个tanh层的简单结构,如图2所示.LSTM的基本结构与基础ANN保持一致,但是重复部分模块包含具有4个网络层的特殊结构,并且利用一种新的方式与其他部分实现交互^[12].因此,解决了RNN的长期依赖以及梯度消失问题^[13].

LSTM的网络结构如图3所示,其中包含输入门、遗忘门、输出门3个逻辑单元,输入门控制记忆单元中当前输入的状态;遗忘门对前一个记忆单元处理结果进行筛选保留;输出门控制记忆单元的输出状态^[14].

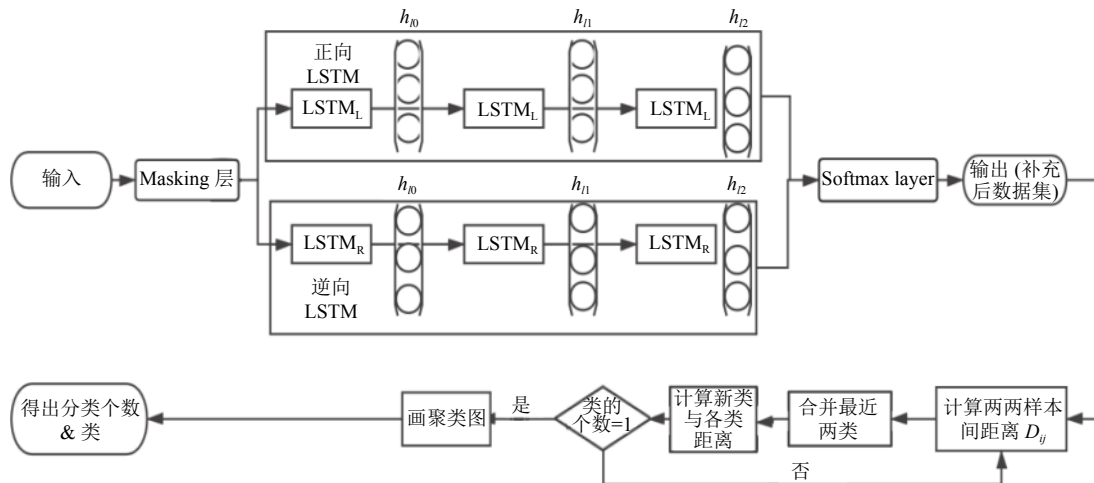


图1 基于 BiLSTM 改进聚类模型

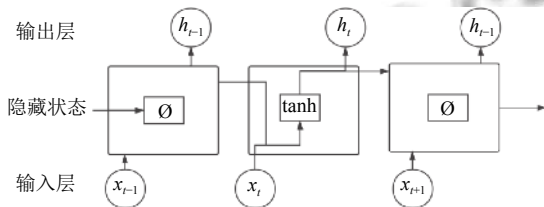


图2 标准 RNN 结构图

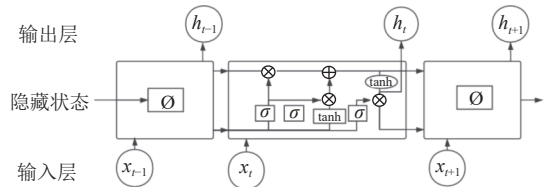


图3 LSTM 神经网络结构图

LSTM 的初始步骤在于判断细胞状态中信息的去留. 这一判断由遗忘门层进行.

遗忘门的状态更新公式为:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

它接收 h_{t-1} 和 x_t , 对于细胞状态 C_{t-1} 中的每部分的输出值都限制在 0 到 1 之间. 接受程度跟随 0 到 1 数值大小递增, 0 表示完全不接受, 1 表示完全接受.

下一步的目的在于判断细胞状态中需要更新的信息并建立新信息. 通过输入门层决定需要被更新的数据. 然后, 通过一个 \tanh 形网络层创建一个新的备选值向量 C_t , 可以用来添加到细胞状态.

输入门状态更新公式:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

通过下述公式对细胞状态的更新:

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t$$

最后, 确定输出值. 首先, 确定细胞状态中的输出部分. 然后, 我们把细胞状态输入 \tanh 与 s 形网络层的输出值相乘, 实现输出.

输出门状态更新公式如下:

$$o_t = \sigma(W_o [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t \times \tanh(C_t)$$

对于 LSTM 来说, 只能学习数据对之前数据的依赖性, 无法充分考虑对之后数据的依赖性^[15]. 在面对数据缺失问题时, 需要结合数据缺失前后数据来共同决定当前缺失数据补充. 图 3 为 BiLSTM 模型图.

BiLSTM 可以通过双向计算, 不仅能考虑到缺失数据之前的监测数据记录, 还能捕捉到后续监测数据. 将前向 LSTM 模型和后向 LSTM 模型构成 BiLSTM 模型来学习双向监测数据信息.

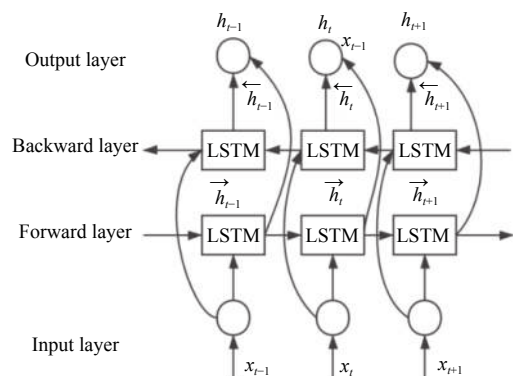


图4 BiLSTM 神经网络结构图

1.3 聚类分析模块

本文采用聚类中的适用于少量特征、少量点位的凝聚层次聚类法(以下简称为层次聚类法)对没有预先处理的微子站点监测数据进行分类^[16-19]。如图4所示,每个样本先自成一类,然后按距离准则逐步合并,减少类数。首先,将微子站分为*n*类(每个微子站为一类),计算类间距离,将类间距最为接近的两类合并为一类,循环计算*n-1*类之间的间距,不断合并各个微子站,直到所有类都完成归类,聚类结束^[20]。聚类分析根据类间距离的计算方法不同,层次聚类法也不同。本文采用最为常用的最小距离法计算类间距离。

最小距离公式:

$$D_{\min} = \min\{D(X_m, X_n)\}$$

其中, $D(X_m, X_n)$ 为*M*类中的某个样本 X_m 和*N*类中的某个样本 X_n 之间的欧式距离。

欧式距离公式:

$$dist(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

层次聚类法示意图如图5所示。

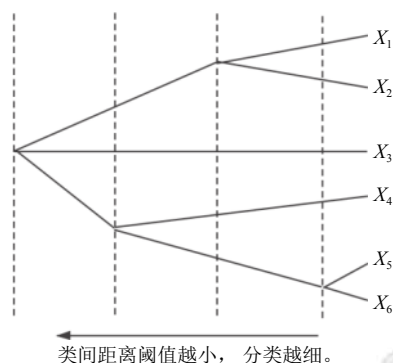


图5 层次聚类示意图

2 实验与结果分析

为验证本文提出的基于 BiLSTM 改进的聚类算法的效果,本实验数据集来源为沈阳市浑南区 18 个微子站 2020 年 1 月-2021 年 8 月的 SO₂、NO₂、O₃、PM₁₀、PM_{2.5}、CO 六种大气污染物的日均监测数据,表 1 为其中某月各站点 6 种污染物日均监测数据。

实验中选取的 18 个监测站点位置如图 6 所示。

2.1 原始数据处理

部分原始数据如图 7 所示,在图 7 中横轴代表的是部分月份各站点的 CO 日均值原始数据组数,纵轴

代表的是 CO 原始数据的浓度值。从图上可以看出部分原始数据存在缺失的情况。因此,在使用此数据集进行模型训练和验证之前需要对原始数据进行数据预处理,补充缺失数据。

表 1 主要污染物浓度监测日均值 (μg/m³)

站点名称	站点编号	SO ₂	NO ₂	PM ₁₀	CO	O ₃	PM _{2.5}
第三热源厂	站点1	24	33	48	1221	52	39
浑南中路	站点2	27	29	42	1300	53	32
新才街	站点3	27	25	42	1221	51	35
仓储街	站点4	25	33	51	1246	52	41
新秀街	站点5	27	24	87	1271	50	65
第二热源厂	站点6	21	26	62	1242	53	47
安姆大厦	站点7	27	29	89	1292	52	65
前榆村	站点8	27	22	109	1229	48	85
金阳大街	站点9	23	29	68	1150	52	53
南屏东路	站点10	27	28	78	1246	52	61
南堤东路	站点11	27	26	44	1246	46	37
银卡东路	站点12	25	27	43	1217	52	37
全运路	站点13	27	25	50	1288	51	32
白塔河二路	站点14	28	36	71	1217	52	57
学城路	站点15	27	33	44	1242	50	36
文汇街	站点16	23	34	35	1246	52	31
高荣路	站点17	22	30	121	1300	55	73
苏王线东段	站点18	26	34	103	1640	53	80



图6 监测站点位置

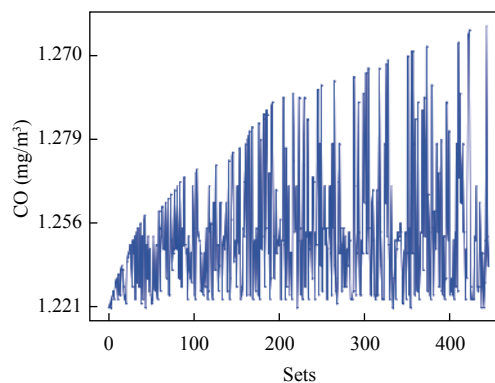


图7 CO 原始数据

同时,数据的标准化处理是数据处理过程中的至关重要的一环,将原始观测数据矩阵中的每类元素,按照某种特定的运算把它变为一个新值,使其更具有可比性.因此,在聚类分析之前,对数据进行预处理,将原始数据映射到(0,1)之间,即均匀化原始数据公式如下:

$$x_{ij}^0 = (x_{ij} - x_{imin}) / (x_{imax} - x_{imin})$$

其中, x_{ij}^0 表示标准、归一化处理后的数据, x_{ij} 表示第 i 种污染物在第 j 个微子站的监测原始数据, x_{imin} 表示的是第 i 种污染物在所有监测点位中的最小浓度值, x_{imax} 表示的是第 i 种污染物在所有监测点位中的最大浓度值.

2.2 实验环境

本文实验所采用的硬件配置为: Intel(R) Core(TM) i5-1035G1 CPU@1.0 GHz 的处理器,内存 DDR4 (16 GB); 软件环境为: 编辑器 PyCharm 2019.3.2, Python 3.7, PyTorch 1.5.

2.3 实验过程

实验过程如图5所示,首先,对数据进行标准化、归一化处理.由于原始监测数据存在缺失,因此,对序列缺失处,补上日期索引并特殊值标记该时间步,之后,Masking层接收经过预处理后的序列,为后续BiLSTM层过滤掉标记为缺失的时间步,为保证监测数据季节完整性以及客观地展示本文方法的补全效果,将2020年1-12月的数据作为训练集,2021年1-4月的数据作为验证集,输入上述的BiLSTM网络模型中,开始训练,设置误差平方和 $E=0.001$; 学习率 $\eta=0.01$. 经过589次迭代训练网络达到要求.最后,将2021年5-8月的数据随机选取300个监测数据,删除其真实值作为缺失数据,并将处理后数据作为测试样本输入已训练好的网络中进行估计,得到缺失数据的估计值.图8为部分缺失数据估计值与真实值的误差曲线图,可以看出BiLSTM神经网络模型结果与实测值变化趋势基本一致,由此可得出该神经网络模型训练结果良好.

为了更好地验证BiLSTM神经网络模型的监测数据插补方法的效果,本文将提出的数据缺失补充模型与以往在空气质量数据缺失情况下常用的均值插补法以及使用单向LSTM神经网络进行对比.为了评价模型描述实验数据的精确度与反映补充值误差的实际情况,本文采用计算简便的均方根误差(RMSE)与具有更强鲁棒性的绝对平均误差(MAE)来共同分析补充

结果和真实测量值之间的偏差.表2是3种模型关于部分站点数据补充结果的评价指标的对比,一般来说当预测值与真实值之间的偏差越小,也就是说MAE和RMSE的值越小,模型的补充效果越好.

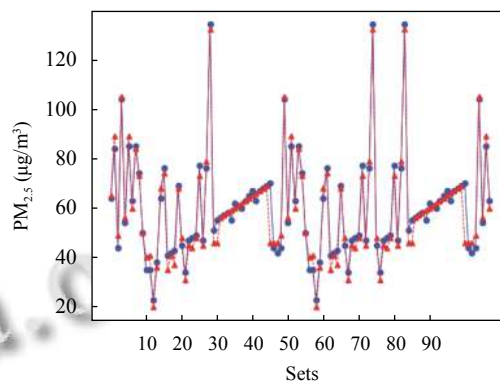


图8 监测数据补充值与真实值对比

表2 不同模型下缺失数据填补实验结果对比

站点名称	污染物	BiLSTM		均值		LSTM	
		RMSE	MAE	RMSE	MAE	RMSE	MAE
浑南中路	SO ₂	18.34	7.90	19.67	10.24	18.57	8.21
	NO ₂	8.3	4.92	12.33	5.22	10.21	5.10
	PM ₁₀	10.3	6.25	12.44	7.66	11.34	6.03
	CO	5.42	4.08	9.42	4.08	7.14	4.22
	O ₃	9.20	5.84	14.42	7.84	10.12	6.32
	PM _{2.5}	14.14	6.45	18.24	8.40	15.68	7.60
新才街	SO ₂	14.35	6.46	20.67	11.21	17.76	7.46
	NO ₂	5.30	3.62	7.33	3.94	6.67	4.02
	PM ₁₀	12.30	7.35	14.54	8.23	15.3	8.15
	CO	7.42	3.58	9.02	4.35	8.44	4.56
	O ₃	11.20	5.61	15.02	8.43	13.06	6.16
	PM _{2.5}	16.14	8.91	20.14	11.02	19.44	9.71
仓储街	SO ₂	10.35	4.76	14.17	6.64	15.47	6.64
	NO ₂	4.83	3.55	7.43	3.94	6.45	4.96
	PM ₁₀	10.34	5.23	14.64	7.11	14.64	6.43
	CO	9.42	3.61	11.02	4.13	10.20	5.01
	O ₃	8.20	3.63	13.02	4.36	11.42	3.46
	PM _{2.5}	14.12	6.64	19.04	8.72	16.56	7.27

由以上分析可以得出本文提出的模型在空气质量数据缺失的补充效果表现更优.进而在此基础上,基于补充数据进行聚类分析,从18个采样点中可选出1、2、3、5、6、8、9、10、13、14共10个优化点,如图9所示.

为验证算法选取的优化点位准确度,对比单独使用原始数据进行层次聚类法选取的优化结果,如图10所示,发现与单独使用聚类分析法选择的1、2、4、

3、13、16、8、5、6、10、14、17等12个点位基本保持一致。

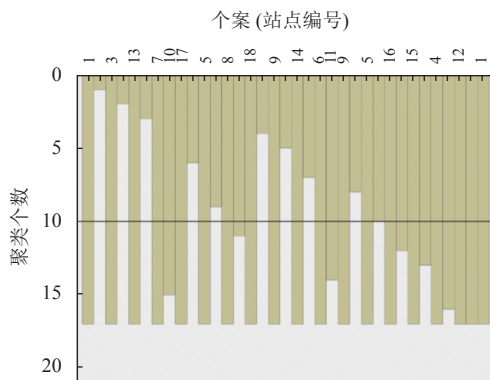


图9 填补数据层次聚类结果图

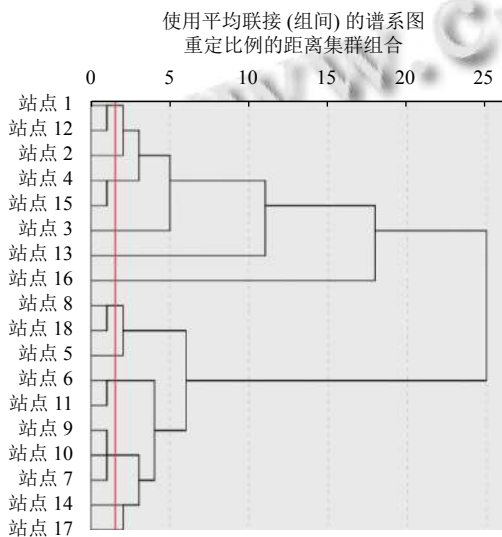


图10 原始数据层次聚类结果图

2.4 优化结果检验

本文通过相关性检验,推断点位优化前后点位的可靠性,通过选取沈阳市浑南区2021年4-8月6种主要污染物浓度日均实测值进行相关性检验,判断优选后的微子站能否客观地反映空气质量。

相关系数公式:

$$r = \frac{\sum_{i=1}^n (X_i - E(X))(Y_i - E(Y))}{\sqrt{\left(E(X^2) - \frac{(E(X))^2}{N}\right)\left(E(Y^2) - \frac{(E(Y))^2}{N}\right)}}$$

其中, X_i, Y_i 指优化前后污染物浓度值。

在给定 $\alpha=0.05$ 时,相关性表如表3。

根据表1所示的相关性检验,表明原点位与优化

后点位的各污染物浓度密切相关,无明显差异性。

为进一步验证优化前后监测数据的一致性,选取部分监测数据采用方差齐性检验-F检验法,对优化前后数据的显著性差异进行检验。在显著性差异不明显的前提下,继续进行t检验,确定两组数据是否具有 consistency。结果如表4所示,可以得出优化前后各污染物浓度之间具有一致性。由此表明优化点位具有较好的代表性,点位选择较为准确,表明优化后的点位可以替代原点位。

表3 各污染物相关性系数表

指标	SO ₂	NO ₂	O ₃	PM ₁₀	PM _{2.5}	CO
$r_{\text{计}}$	0.915	0.932	0.908	0.914	0.935	0.878
$r_{0.05(n-2)}$	0.878	0.878	0.878	0.878	0.878	0.878

表4 一致性检验结果

指标	SO ₂	NO ₂	O ₃	PM ₁₀	PM _{2.5}	CO
$F_{\text{计}}$	1.84	1.45	1.04	1.68	1.31	1.22
$F_{\text{表}}$	2.11	2.11	2.11	2.11	2.11	2.11
$t_{\text{计}}$	0.34	0.23	0.01	0.85	0.56	0.39
$t_{\text{表}}$	1.96	1.96	1.96	1.96	1.96	1.96

2.5 性能评价指标

聚类属于无监督学习中的一种,原始数据不具有标签,需要一些定量的指标来进行评估算法的好坏^[19,21]。根据是否提供样本的标签信息,相关的指标可以分为外部方法与内部方法,内部方法指的是不需要数据的标签,仅仅从聚类效果本身出发,而制定的一些指标^[16]。因为点位优化属于无标签数据,故选择内部方法的2种评价指标:CH指数(Calinski-Harabaz index, CH)、戴维森堡丁指数(Davies-Bouldin index, DBI)作为评价标准,对基于补充数据后数据集与基于原始数据集的聚类效果进行评价。

CH指数,综合考虑了簇间距离和簇内距离,计算公式如下:

$$CH = \left[\frac{\sum_{k=1}^K n_k \|c_k - c\|^2}{K-1} \right] \left[\frac{\sum_{k=1}^K \sum_{i=1}^{n_k} \|d_i - c_k\|^2}{N-K} \right]$$

CH的数值大小与簇内距离成负相关,簇间距离越大,聚类效果越好。

戴维森堡丁指数,公式如下:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} \left(\frac{\text{avg}(C_i) + \text{avg}(C_j)}{d_{\text{cen}}(C_i, C_j)} \right)$$

其中, $\text{avg}(C)$ 表示聚类簇的紧密程度, 公式如下:

$$\text{avg}(C) = \frac{2}{|C|(|C|-1)} \sum_{1 < i < j \leq |C|} \text{dist}(x_i, x_j)$$

计算该聚类簇内样本点的距离, d 表示不同聚类簇中心点之间的距离, 公式如下:

$$d_{\text{cen}}(C_i, C_j) = \text{dist}(u_i, u_j)$$

聚类簇间距离越远, 聚类簇内距离越近, DBI 指数越小, 聚类算法性能越好。

BiLSTM 神经网络填补监测数据与原始数据两种状态下, 进行层次聚类后, 聚类效果评价指标对比如表 5 所示, 可以看出相比较于对原始数据进行聚类操作, 本文所采用的基于 BiLSTM 改进的聚类算法聚类效果有所提升。

表 5 评价指标对比

方法	CH	DBI
凝聚层次聚类法	937	0.756
凝聚层次聚类法+BiLSTM	1198	0.323

2.6 结果分析

根据 BiLSTM 神经网络补充数据的改进聚类算法从 18 个监测点位中选出了 10 个监测点位, 与在原始数据上使用层次聚类法选出的 12 个点位基本保持一致, 对比单独使用聚类算法选出的 12 个监测点位。此模型选择的点位更少, 并且聚类效果更好, 选择的点位代表性与准确性均得到提升, 为之后监测点位布设与优化提供一个更加高效准确的模型。由表 3 的计算结果可以推断本文提出的方法选择的点位是准确且具有代表性的。

3 结论

本文针对空气质量监测点位优化的最终目的, 通过层次聚类法对 BiLSTM 神经网络补充后的微子站监测数据进行聚类处理, 在选择更加少量点位的基础上, 提高了点位选择准确度, 提出了一种基于 BiLSTM 神经网络改进聚类的点位优化算法。

参考文献

- 赵晓亮, 齐庆杰, 赵东洋, 等. 阜新市空气监测点位优化的聚类分析. 地球与环境, 2015, 43(3): 350-355.
- 王陆平. 大气质量评价模型和监测点位优化研究 [硕士学位论文]. 西安: 西安科技大学, 2017.

- 刘潘炜, 郑君瑜, 李志成, 等. 区域空气质量监测网络优化布点方法研究. 中国环境科学, 2010, 30(7): 907-913.
- 徐明德, 陈武. 模糊物元法在优选大气监测点中的应用. 测试技术学报, 2007, 21(5): 391-395.
- 杜增荣. 空气质量监测点位布设现状及优化布点研究. 山西化工, 2020, 40(4): 208-211.
- 姜林, 徐婷婷, 邹长武. BP 网络应用于大气环境监测点的优选. 环境科学与技术, 2003, 26(5): 19-20, 65.
- 李董, 迟家俊, 相博, 等. 基于 SMOTE 和 KNN 的石油数据缺失填充算法. 数学的实践与认识, 2019, 49(17): 187-195.
- 袁兆祥, 余春生. 基于 DBSCAN 聚类的电力工程数据完整性分析. 沈阳工业大学学报, 2019, 41(3): 246-250. [doi: 10.7688/j.issn.1000-1646.2019.03.02]
- 杨挺, 何周泽, 赵东艳, 等. 基于 FSOM 神经网络的电能质量数据缺失修复算法. 电网技术, 2020, 44(5): 1941-1949.
- 柯昊, 李天, 周悦, 等. 数据缺失时基于 BP 神经网络的作者重名辨识研究. 情报学报, 2018, 37(6): 600-609. [doi: 10.3772/j.issn.1000-0135.2018.06.005]
- 宋维, 高超, 赵玥, 等. 基于 LSTM 的活立木茎干水分缺失数据填补方法. 林业科学, 2020, 56(2): 134-141.
- 杨青, 王晨蔚. 基于深度学习 LSTM 神经网络的全球股票指数预测研究. 统计研究, 2019, 36(3): 65-77.
- 荣光辉, 黄震华. 基于深度学习的问答匹配方法. 计算机应用, 2017, 37(10): 2861-2865.
- 王毅, 谢娟, 成颖. 结合 LSTM 和 CNN 混合架构的深度学习神经网络语言模型. 情报学报, 2018, 37(2): 194-205.
- 范秀琴, 喻洪流, 李素姣. 基于 BiLSTM 的表面肌电图手势识别算法. 生物医学工程学进展, 2021, 42(2): 92-96.
- 乔端瑞. 基于 K-means 算法及层次聚类算法的研究与应用 [硕士学位论文]. 长春: 吉林大学, 2016.
- 吕言成. 基于神经网络降维聚类的水质点位优化系统设计与实现 [硕士学位论文]. 沈阳: 中国科学院大学 (中国科学院沈阳计算技术研究所), 2021.
- 刘涵, 王宇, 马琰. 改进聚类的深度神经网络压缩实现方法. 控制理论与应用, 2019, 36(7): 1130-1136.
- 汪庆森. 基于目标函数的模糊聚类新算法及其应用研究 [博士学位论文]. 镇江: 江苏大学, 2014.
- Insung Y, Tran HQ, Oun HS. Classification of network packets using hierarchical clustering. Journal of the Korea Internet of Things Society, 2017, 3(1): 9-11.
- Aghazadeh M, Gharehchopogh FS. A new hybrid model of multi-layer perceptron artificial neural network and genetic algorithms in Web design management based on CMS. Journal of AI and Data Mining, 2018, 6(2): 409-415.