

彭旭, 饶元, 乔焰. 基于宽度卷积神经网络的异常农情数据检测方法 [J]. 华南农业大学学报, 2022, 43(2): 113-121.
PENG Xu, RAO Yuan, QIAO Yan. Detection method of abnormal agricultural data based on broad convolution neural network[J]. Journal of South China Agricultural University, 2022, 43(2): 113-121.

基于宽度卷积神经网络的异常农情数据检测方法

彭旭¹, 饶元¹, 乔焰²

(1 安徽农业大学 信息与计算机学院, 安徽 合肥 230036; 2 合肥工业大学 计算机与信息学院, 安徽 合肥 230001)

摘要:【目的】为准确有效地检测农业物联网的感知数据异常, 提出了基于宽度卷积神经网络的异常农情数据检测方法, 为实现农业物联网数据高质量感知提供参考。【方法】首先将标准化后的农情数据编码为极坐标表示, 通过滑动窗口机制划分子集, 接着将每个子集数据重构为矩阵, 最后设计并训练宽度卷积神经网络模型用于异常检测, 采用养殖场环境监测数据进行试验。【结果】构建的滑动窗口机制可提升异常数据检测能力, 缩短检测时间。所设计的宽度卷积神经网络对空气温湿度、土壤温湿度等数据中所存在的异常检测准确率均超过 97.5%, 优于 SVM、RF 和 CNN 模型 1.69%、2.76% 和 3.05%; $F1$ 值均在 0.985 以上, 优于 SVM、RF 和 CNN 模型 0.009 3、0.014 9 和 0.016 3; 且在处理波动性较大的空气、土壤温湿度数据时性能优势更为明显, 准确率和 $F1$ 值分别提高了 3.61%~5.98% 和 0.018 8~0.031 0。此外, 该方法模型检测耗时较短, 仅为传统 CNN 模型的 1/6~1/7, 并且比 SVM 和 RF 模型使用更少的超参数。【结论】所建立的数据编码、子集划分和重构方法与宽度卷积神经网络模型对异常农情数据有较好的检测效果。

关键词: 农业物联网; 传感器; 宽度卷积神经网络; 异常检测
中图分类号: S24; TP391 **文献标志码:** A **文章编号:** 1001-411X(2022)02-0113-09

Detection method of abnormal agricultural data based on broad convolution neural network

PENG Xu¹, RAO Yuan¹, QIAO Yan²

(1 School of Information & Computer, Anhui Agricultural University, Hefei 230036, China; 2 School of Computer Science and Information Engineering, Hefei University of Technology, Hefei 230001, China)

Abstract: 【Objective】In order to accurately and effectively detect data anomalies from agricultural internet of things, one detection method of abnormal agricultural data based on broad convolution neural network (BCNN) was proposed for providing a reference for achieving high-quality data collection in agricultural internet of things. 【Method】Firstly, the standardized agricultural data were encoded as polar coordinates, and then they were divided into subsets by sliding window mechanism. Subsequently, the data were reconstructed as matrix format. Finally, the BCNN was designed and trained for conducting anomaly detection. The experiment was conducted using the data monitored in the culturing farm environment. 【Result】The sliding window mechanism could improve the detection ability and reduce the time consumption. The accuracy and $F1$ score of

收稿日期: 2021-03-30 网络首发时间: 2021-09-01 14:14:06
网络首发地址: <https://kns.cnki.net/kcms/detail/44.1110.S.20210901.0947.002.html>
作者简介: 彭旭, 硕士研究生, 主要从事计算机应用技术研究, E-mail: ernstpeng@163.com; 通信作者: 饶元, 教授, 博士, 主要从事计算机应用技术研究, E-mail: raoyuan@ahau.edu.cn
基金项目: 农业农村部农业国际合作项目 (125A0607); 安徽省自然科学基金 (2008085MF203); 安徽省重点研究和开发计划面上攻关项目 (201904a06020056); 安徽农业大学省级大学生创新创业训练计划项目 (S202010364228)

the designed BCNN in the datasets of air temperature and humidity, soil temperature and humidity were more than 97.5% and 0.985 respectively, which on average outperformed SVM, RF and CNN with the increase of 1.69%, 2.76%, 3.05% and 0.009 3, 0.0149, 0.0163, respectively. In particular, while handling the air and soil temperature and humidity data with high fluctuation , the gain in accuracy and *F1* score ranged 3.61%–5.98% and 0.018 8–0.031 0, respectively. In addition, the proposed BCNN model has less time consumption of anomaly detection, only 1/6 to 1/7 of classical CNN model, and as well as with less hyperparameter. 【Conclusion】 The proposed data preprocessing (data coding, subset partition and reconstruction) method and BCNN model exhibit better performance on abnormal agricultural data.

Key words: Agricultural internet of things; Sensor; Broad convolution neural network; Anomaly detection

精准农业是推动农业现代化与信息化发展的关键，其重要的农情数据采集阶段离不开农业物联网 (Internet of things in agriculture) 的支持^[1]。现阶段，农业物联网已在农业生产中得到广泛应用，复杂的传感器网络按时间序列连续地采集温度、湿度、CO₂ 和 NH₃ 浓度等大量的环境及作物生长信息。分析和评估采集的信息可为农业精准决策提供保障。因此，农业物联网数据的有效挖掘和利用已成为农业信息化领域的研究热点^[2-3]。然而，由于农业作业环境往往较为恶劣，且受制造工艺和网络传输的限制，导致采集到的数据中不可避免地存在异常^[4]，这些异常可能干扰数据分析并影响农业精准决策。

费欢等^[5]以传感器网络多模态数据流之间的相干性为理论基础，利用多维数据和滑动窗口模型对异常数据及其来源进行检测和评估。Zhao 等^[6]提出一种基于多分类器集成的漂移补偿监督学习算法，利用支持向量机 (Support vector machine, SVM) 和改进的长短期记忆 (Long short term memory, LSTM) 构建多分类器模型，通过归一化和加权策略，在每次分类过程中去除精确度最低的基分类器，使模型适应传感器漂移，有效地提高了传感器漂移分类性能。Wang 等^[7]利用传感器之间的数据相关性，使用信号空间投影和卡尔曼滤波器实现传感器漂移的盲校准。高鹏等^[8]利用 LSTM 和广义回归神经网络 (Generalized regression neural network, GRNN) 模型对土壤墒情进行建模，实现对柑橘土壤含水量和土壤电导率的预测。

在传统的机器学习任务中，需要用多种信号处理算法检测数据流中的异常，导致效率较低，在数据分析过程中易出现过拟合或欠拟合现象。此外，大多数异常检测算法泛化能力较弱，不适用于其他类型的异常检测。文献 [6] 中采用 SVM 和 LSTM 的结合模型，获得了理想精确度，但增加了时间复

杂度。文献 [7] 所提信号空间投影和卡尔曼滤波器方法，利用传感器之间的相关性解决传感器漂移问题，但传感器难以密集部署，模型应用场景狭小、泛化能力较弱。

近年来，随着硬件计算能力的显著提升，人工智能 (Artificial intelligence) 和深度学习 (Deep learning) 技术已广泛应用于图像识别^[9]、自然语言处理^[10]、推荐系统^[11] 和音频识别^[12] 领域。深度结构的神经网络在传感器数据异常检测领域也得到广泛应用，Bao 等^[13]提出了一种基于计算机视觉和深度学习的传感器数据异常检测方法，该方法将时间序列信号转换为图像并保存在灰度图像中，通过训练两隐藏层堆叠的自动编码深度神经网络来自动检测未标签的数据。Tang 等^[14]提出了一种基于卷积神经网络的传感器数据异常检测方法，在图片中融合数据的时域和频域特征，并采用卷积神经网络对图片进行分类，但时域和频域可视化方法无法保存数据的时间数值依赖性。Liu 等^[15]提出一种动态宽度增量学习算法，通过深度叠加宽度学习系统，解决数据回归和图像分类问题，但在数据预处理中，未采用有效方法增强数据时间依赖性。Wang 等^[16]提出一种格拉姆角求和域/差分域和马尔科夫过渡域的时间序列数据可视化方法，并采用平铺卷积神经网络学习图像特征，实现时间序列数据分类，但当原始时间序列过长时，重构后的矩阵规模大，增加了运算时间消耗。

针对以上研究现状，本文设计了一种宽度卷积神经网络 (Broad convolution neural network, BCNN) 的农情数据异常检测方法，并通过滑动窗口机制解决单次输入模型数据量大引起的模型检测耗时长和准确率不足等问题。研究数据预处理与重构方法、滑动窗口尺寸对异常检测性能的影响，对比分析不同异常检测模型的性能差异，以期精准农业的数据高质量采集提供参考。

1 材料与方法

1.1 试验材料

试验数据采用位于安徽省长丰县双墩镇的合肥安谷农业有限公司羊圈监测点传感器的空气湿度、空气温度、土壤湿度和土壤温度数据,时间覆盖范围从2019年12月1日至2019年12月31日,采样间隔3 min,共11 355条观测数据。为便于开展检测性能对比分析,在数据中引入一组随机故障率

(10%~40%)的异常点^[17]。为系统评估所提方法的性能,采用80%的数据集作为训练集、20%的数据集作为测试集。数据集如图1所示,4种数据的周期性和趋势特征存在差异。空气湿度数据值域波动较大,周期性特征不明显。空气温度数据呈周期性变化,波动较小。土壤湿度数据具有明显的周期性特征,值域范围较小。土壤温度数据周期性特征与空气温度数据相似,且值域波动更小。

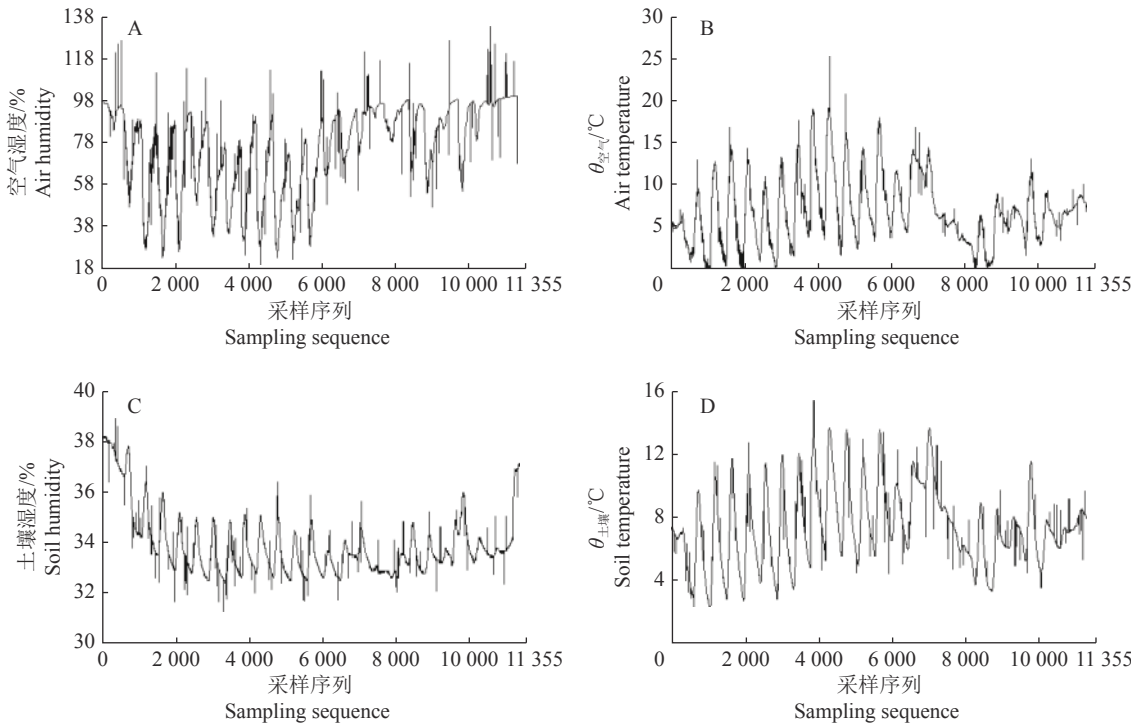


图1 试验数据集
Fig.1 Experimental dataset

1.2 异常检测框架

基于宽度学习的农情数据异常检测框架图如图2所示。传感器端负责采集数据并发送至数据中心,数据中心负责存储数据、训练模型和异常检测。模型训练过程如下:部署传感器并采集数据,将数据归一化,基于数据周期和采样间隔等参数确定滑动窗口尺寸 l ,编码为极坐标表示,并基于滑动窗口尺寸 l 划分子集,采用格拉姆角求和域 (Gramian summation angular field, GASF) 方法重构为矩阵,最后训练得到 BCNN 数据异常检测模型。

在异常数据实时检测过程中,数据中心对采集的最新数据进行归一化处理,编码为极坐标并重构矩阵,最后输入模型检测异常。若模型判断为正常,则将其保存为历史数据,否则,校准异常并保存。校准异常可采用回归模型预测、均值替换等方法^[18]。

1.3 数据预处理

设传感器节点在某时刻共采集到 n 个时间序列数据 $X = \{x_1, x_2, \dots, x_n\}$,为降低采样数据取值范围对模型训练效果的影响,将数据归一化至 $[-1,1]$ 区间内,表示为:

$$x'_i = \frac{(x_i - X_{\max}) + (x_i - X_{\min})}{X_{\max} - X_{\min}}, i \in [1, n], x'_i \in X', \quad (1)$$

式中, x_i 表示原始时间序列数据, X' 表示归一化后的时间序列数据, $X_{\max}$ 和 $X_{\min}$ 分别表示原始时间序列中的最大值和最小值, x'_i 表示归一化后的数据。归一化将原始数据映射至 $[-1,1]$ 区间内,将有量纲数据转化为无量纲形式,降低不同量纲对模型的影响,并保留原始数据特征,使不同的数据集处于相同数量级,便于对比分析。

模型处理一维时间序列时,无法有效保持数据

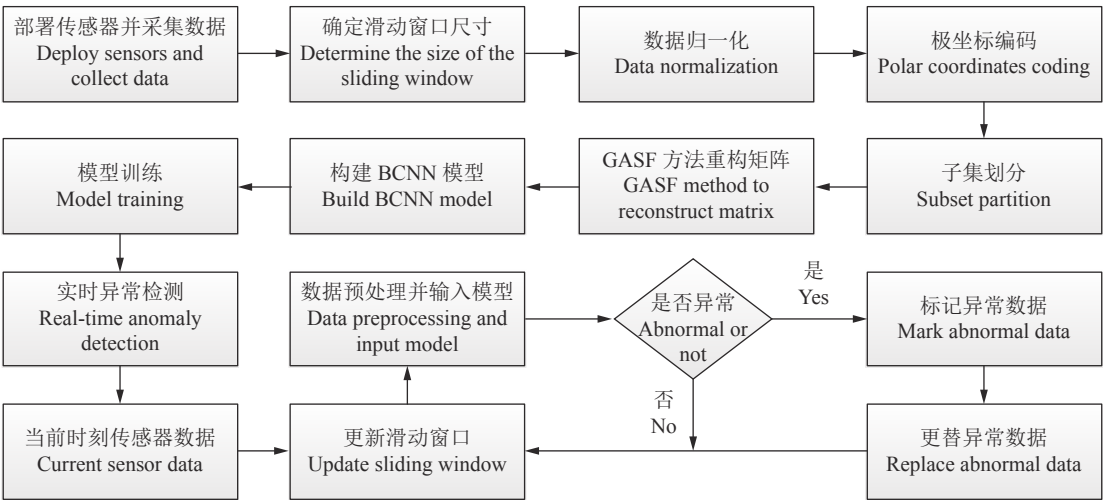


图 2 基于 BCNN 的传感器数据异常检测框架

Fig. 2 Anomaly detection framework of sensor data based on BCNN

周期性时间依赖,无法有效提取部分数据之间的相关性,且对于具有不同周期的时间序列数据,模型泛化能力较差。为此,本文通过将时间序列数据编码为极坐标并重构为矩阵,实现模型对时间序列数据时间依赖性的保持和泛化性的增强。

具体的,将归一化后的数据编码为极坐标^[16],时间序列数据的数值和时间戳分别编码为极角和极径,表示为:

$$\theta = \arccos x'_i, x'_i \in X'$$
$$r = \frac{t_i}{n}, t_i \in [1, 2, 3, \dots, n],$$

(2)

式中, t_i 表示时间序列数据的时间戳, n 表示时间序列数据数量, θ 和 r 分别表示编码后数据的极径和极角。基于极坐标的编码方法是预处理和分析时间序列数据的一种新方法。反余弦函数在 $[-1, 1]$ 区间内是单调的,因此,式 (2) 在时间序列数据和极坐标之间构建了一个双射。给定一组时间序列数据,可产生唯一的映射关系,且相对笛卡尔坐标系,极坐标可保留绝对时间关系。

滑动窗口控制单次输入模型的数据量,是影响异常检测模型性能的重要参数,选取合适的滑动窗口尺寸可有效提高模型准确率和降低时间消耗。本文设计了基于数据特征的滑动窗口尺寸 l 的计算方法:

$$l = \left(\frac{\Delta t}{\beta} + \ln S \right) T e^{-\beta \log_2 \sqrt{T/\Delta t}},$$

(3)

式中, β 为支持度衰减因子, Δt 为农业物联网传感器数据采样间隔, T 为农情数据特征周期, S 为农情数据的标准差。滑动窗口尺寸对模型准确率和效率存在影响,窗口尺寸过大会增加模型检测耗

时,过小则会导致模型无法有效提取数据特征,从而降低准确率。式中 $T/\Delta t$ 控制滑动窗口选择函数上升幅度和衰减速度, $\Delta t/\beta + \ln S$ 平衡采样间隔和数据特征的影响,可满足模型异常检测准确率和效率需求。

根据滑动窗口尺寸 l ,将编码后的数据分为若干个子集 $\{\theta'_1, \dots, \theta'_l\}, \{\theta'_{l+1}, \dots, \theta'_{2l}\}, \dots$,并将每个子集数据采用 GASF 方法重构为矩阵,采用类热力图形式可视化:

$$\mathbf{G} = \begin{pmatrix} \cos(\theta'_1 + \theta'_1) & \cdots & \cos(\theta'_1 + \theta'_l) \\ \vdots & \ddots & \vdots \\ \cos(\theta'_l + \theta'_1) & \cdots & \cos(\theta'_l + \theta'_l) \end{pmatrix},$$

(4)

式中, $\mathbf{G}$ 为重构后的矩阵, θ'_1 和 θ'_l 为编码为极坐标并划分子集后的数据。

通过式 (2) 将归一化后的数据编码为极坐标,采用式 (3) 重构为矩阵。时间序列数据编码为极坐标后仍保留数据连续性,重构后的矩阵随着位置的变化,时间会增加,可有效保存时间依赖性。此外,矩阵也包含了时间相关性,其中的 $G_{i,j}$ 元素叠加了原始数据中第 i 和第 j 个数据相关性,主对角线包含了原始数据信息。因此,可采用神经网络提取高维数据特征。

为展示完整的数据预处理方法,以土壤湿度数据为例,完整流程如图 3 所示。首先,采用式 (1) 将原始土壤湿度数据归一化至 $[-1, 1]$ 区间内,并通过式 (2) 编码为极坐标表示 (图 3B),极坐标数据保留时间依赖性,原始异常数据在极坐标系中仍存在离群特征 (即极坐标图中存在显著的异常点),接着,采用式 (3) 将极坐标数据划分子集,并通过式 (4) 重构矩阵。图 3C 展示了部分矩阵,异常数据可在

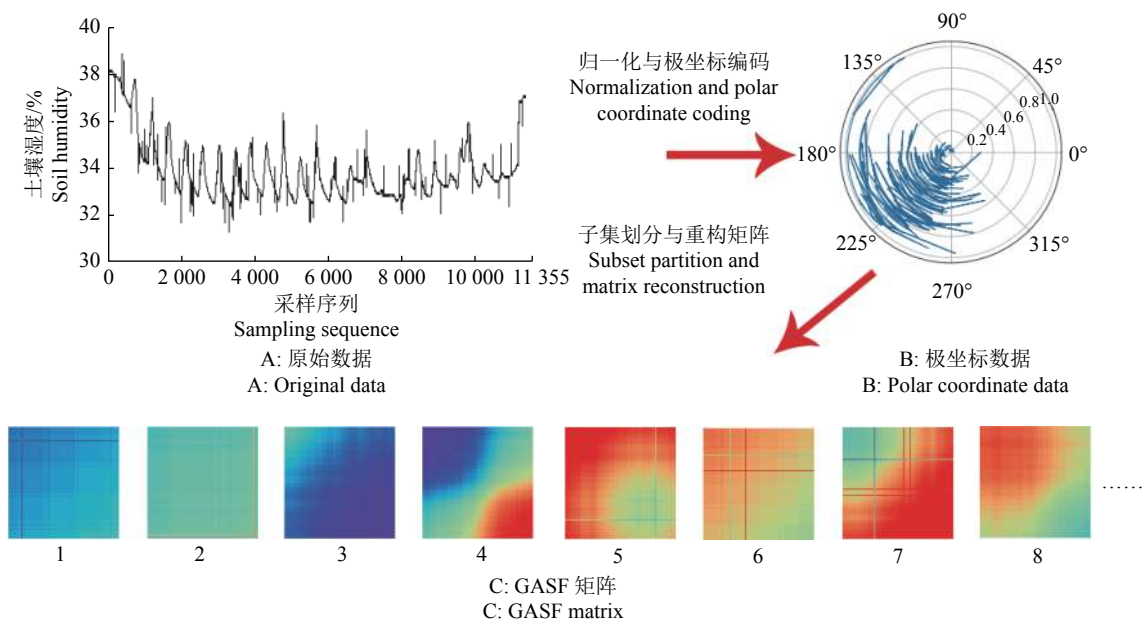


图 3 土壤湿度数据预处理流程

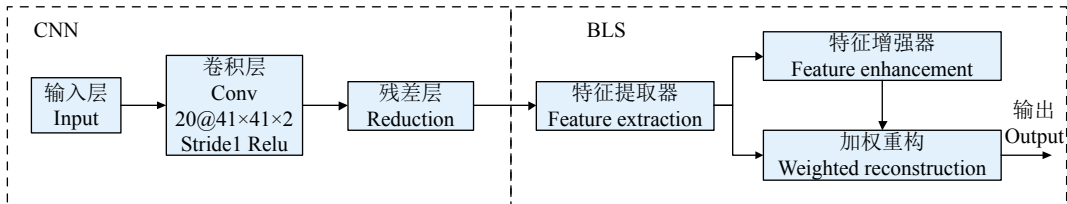
Fig. 3 Soil humidity data preprocessing process

GASF 矩阵中显著表示, 矩阵中出现颜色突变即表示该点数据存在异常, 如矩阵 1、2、5、6 和 7, 矩阵中颜色连续变化即表示无异常, 如矩阵 3、4 和 8。预处理后的数据特征更为显著, 可采用模型更有效地检测异常数据。

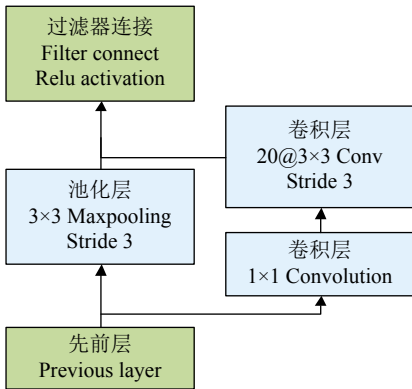
1.4 宽度卷积神经网络框架

设计的宽度卷积神经网络结构如图 4 所示。模型采用宽度学习系统强化传统 CNN 模型^[19]。BCNN

的第 1 层 (Layer 1) 是输入层, 输入矩阵维度为 $k \times k$, 其中 k 表示滑动窗口尺寸。第 2 层 (Layer 2) 是卷积层, 其中卷积核 (Filter) 数量为 20, 每个卷积核尺寸为 41×41 , 步长 (Stride) 为 1, 采用 Relu 激活函数。第 3 层 (Layer 3) 是 Reduction 层, 其中包含 1 个池化模块和 2 个卷积模块, 该层采用残差连接。池化模块中, 池大小 (Pool size) 为 3×3 , 步长为 3, 池化方式为最大池化 (Maxpooling)。卷积模块中,



A: BCNN 结构
A: The structure of BCNN



B: Reduction 模块结构
B: Reduction module structure

图 4 宽度卷积神经网络结构

Fig. 4 The structure of broad convolution neural network

第 1 层卷积采用 1×1 的卷积核, 第 2 层卷积采用 20 个 3×3 大小的卷积核, 步长为 3。卷积模块和池化模块输出的特征图按其深度连接, 并采用 Relu 函数激活, 被平坦化为一维向量输入宽度学习系统。第 4 层 (Layer 4) 采用特征提取器代替传统 CNN 模型中的 Softmax 分类器, 其中包含 1 000 个特征节点。第 5 层 (Layer 5) 为特征增强器, 其中包含 1 000 个特征增强节点, 第 4 层中特征提取器的输出被特征增强器随机映射输出。最后将特征提取器和特征强化器的输出连接, 并加权重构, 得到整个模型的输出。

卷积可通过连续地变换提取数据的复杂特征。在卷积层中, 利用卷积核对输入数据进行卷积:

$$x_i^l = \text{Relu} \left(\sum_{n=1}^N x_n^{l-1} f_{n,i}^l \right), \quad (5)$$

式中, x_i^l 表示第 l 层中图片的 i 通道, $f_{n,i}^l$ 表示位于第 l 层中第 n 个卷积核的 i 通道, N 表示 $l-1$ 层中通道总数, Relu() 表示激活函数。

池化是一种降采样形式。它将图像划分为若干个区域, 每个区域独立地按池化方式输出, 减小了数据空间大小, 参数数量和模型计算量也随之下降低, 在一定程度上防止过拟合的发生。本文采用最大池化方式, 表示为:

$$a_{i,n}^l = \max(a_{i,n}^{l-1}, \dots, a_{i,m}^{l-1}), \quad (6)$$

式中, $a_{i,n}^l$ 表示网络 l 层中图片 i 通道的第 n 个位置, $a_{i,n}^{l-1}, \dots, a_{i,m}^{l-1}$ 表示网络 $l-1$ 层中图片 i 通道的池化区域对应的第 n 到 m 个位置。

宽度学习系统^[15]是一种通过随机方法将特征扩展到广阔空间的模型, 其主要由特征节点和增强节点组成。特征节点将输入随机映射, 映射特征由增强节点随机扩展, 将最后所有特征加权连接输出。

残差层的输出被平坦化为一维向量 X 并输入至特征节点, 其表示为:

$$Z_i = \phi(XW_{z_i} + \beta_{z_i}), i = 1, 2, \dots, n, \quad (7)$$

式中, 权重矩阵 W_{z_i} 和偏置矩阵 β_{z_i} 随机生成, $\phi()$ 表示激活函数, Z_i 表示第 i 个特征节点的输出, 所有特征节点的集合记为 $Z^n = [Z_1, \dots, Z_n]$, 其中 n 为特征节点数量。类似的, 特征增强节点可表示为:

$$H_j = \xi(Z^n W_{h_j} + \beta_{h_j}), j = 1, 2, \dots, m, \quad (8)$$

式中, 权重矩阵 W_{h_j} 和偏置矩阵 β_{h_j} 随机生成, $\xi()$ 表示激活函数, H_j 表示第 j 个特征增强节点的输出, 所有特征节点的集合记为 $H^m = [H_1, \dots, H_m]$, 其中

m 为特征增强节点数量。

将特征节点和特征增强节点连接加权重构, 输出可表示为:

$$Y = [Z_1, \dots, Z_n | H_1, \dots, H_m] W = [Z^n | H^m] W, \quad (9)$$

式中, 权重矩阵 W 可通过伪逆法求解, 可表示为:

$$W = [Z^n | H^m]^\dagger Y = (A^m)^\dagger Y, \quad (10)$$

式中, $A^m = [Z^n | H^m]$, $(A^m)^\dagger$ 表示 A^m 的伪逆, 可通过极限方法求得:

$$A^\dagger = \lim_{\lambda \rightarrow 0} (\lambda I + A A^T)^{-1} A^T, \quad (11)$$

式中, A 表示特征节点和增强节点输出组合成的扩展矩阵, λ 表示对权重矩阵 W 的约束, I 表示单位矩阵, A^T 表示转置矩阵。

1.5 检测性能评价指标

本文采用准确率 (Accuracy)、F1 值和模型检测时间 (Time) 等指标评估模型的异常检测性能, 其表示为:

$$\text{准确率} = \frac{T_p + T_n}{T_p + T_n + F_p + F_n}, \quad (12)$$

$$\text{精确度} = \frac{T_p}{T_p + F_p}, \quad (13)$$

$$\text{召回率} = \frac{T_p}{T_p + F_n}, \quad (14)$$

$$F1 \text{ 值} = \frac{2 \times \text{精确度} \times \text{召回率}}{\text{精确度} + \text{召回率}}, \quad (15)$$

式中, T_p 表示正确分类至目标类的数量, F_p 表示其他类错误分类至目标类的数量, T_n 表示其他类正确分类的数量, F_n 表示目标类错误分类至其他类的数量。

准确率是正确分类的数据与数据集数据总量的比值, 反映模型的正确检测能力。精确度 (Precision) 和召回率 (Recall) 分别是正确分类为异常数据的个数与所有分类的异常数据、实际异常数据的比值。F1 值同时兼顾了精确度和召回率。除异常检测能力外, 模型检测时间也是评估模型是否符合实际应用的重要参数。

2 试验与分析

2.1 试验环境与参数

试验采用 Windows 10 操作系统、Intel Core i5 CPU、8GB 内存的台式计算机, Python3.6 编程语言和 Tensor Flow2.0.0 框架用于异常检测模型搭建和应用。主要研究滑动窗口尺寸对模型异常检测性能

的影响及模型对不同特征数据的异常检测灵敏性,以甄选最优的异常检测方法。在此基础上,对比分析不同异常检测模型性能。

试验将 BCNN 与当前使用较为广泛的 SVM、随机森林 (Random forest, RF) 和卷积神经网络方法进行比较。SVM 模型^[20-21]核函数采用径向基函数 (Radial basis function, RBF), 核函数参数 ε 为 0.2, 正则化参数 C 为 1; RF 模型^[22]聚类数 $c=2$, 模糊程度系数 $m=3$, 卷积神经网络方法从文献 [14] 中引用。BCNN 模型的特征节点和特征增强节点数量均为 1 000。对于滑动窗口, 传感器监测数据特征周期 T 为 24 h, 支持度衰减因子 β 取 0.5^[23], 采样间隔 Δt 为 3 min, 空气湿度、空气温度、土壤湿度和土壤温度的滑动窗口尺寸分别为 105、88、76 和 91。为控制模型参数随机初始化和训练引起的随机性, 重复 5 次试验, 以平均数进行结果分析。

2.2 滑动窗口尺寸对检测性能的影响

为评估所提滑动窗口选择方法的有效性, 以 BCNN 模型为例, 采用窗口尺寸 60~120 的数据集进行试验, 结果如图 5 所示。空气湿度数据集中, 异常检测准确率在窗口尺寸 [60,100] 区间内呈上升趋势, 在窗口尺寸 100 时达到峰值; 空气温度数据集中, 窗口尺寸在 [60,90] 区间时, 异常检测准确率随窗口尺寸增大而增加, 窗口尺寸在 90 时准确率表现最佳, 窗口尺寸大于 90 时准确率无显著变化; 土壤湿度数据集中, 在窗口尺寸 [60,70] 区间内准确率呈增加趋势, 在 [70,120] 区间内准确率趋于稳定; 土壤温度数据集中, 在窗口尺寸 [90,100] 时准确率最高。值得注意的是, 土壤湿度数据集呈最高准确率的窗口尺寸远小于其他数据集, 这是由于其数据波动不显著, 模型仅需少量数据即可学习其特征。试验所得出的最佳滑动窗口尺寸, 与式 (3) 得出的窗口尺寸基本相符。因此, 基于数据特征的滑动窗口

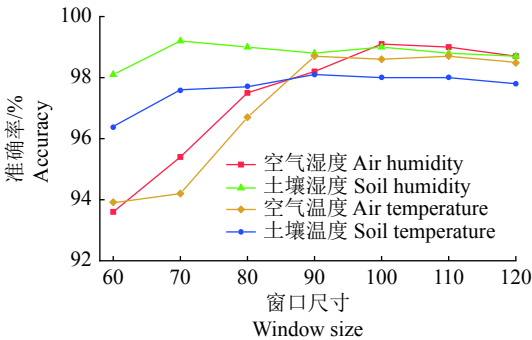


图 5 不同滑动窗口尺寸与数据集的异常检测准确率
Fig. 5 Anomaly detection accuracy of different sliding window sizes and datasets

选择方法可有效地提高模型异常检测准确率。

模型检测耗时随滑动窗口尺寸变化如图 6 所示。空气湿度数据集中, 滑动窗口尺寸在 [90,100] 区间时, 检测耗时最短。空气温度数据集中, 在窗口尺寸 [80,100] 区间时检测耗时最短, 处于理想状态。土壤湿度数据集中, 当窗口尺寸变化时, 模型检测耗时变化不明显, 窗口尺寸 80 左右时检测耗时最短。土壤温度数据集中, 在窗口尺寸 [80,90] 区间内, 检测耗时较短, 处于理想水平, 其中窗口尺寸 90 时最优。总体上, 窗口尺寸和检测耗时呈二次相关。波动性较强的数据 (如空气湿度), 窗口尺寸变化时, 检测耗时波动更大; 而波动小、较为平稳的数据 (如土壤湿度), 改变滑动窗口尺寸对检测耗时影响较小。

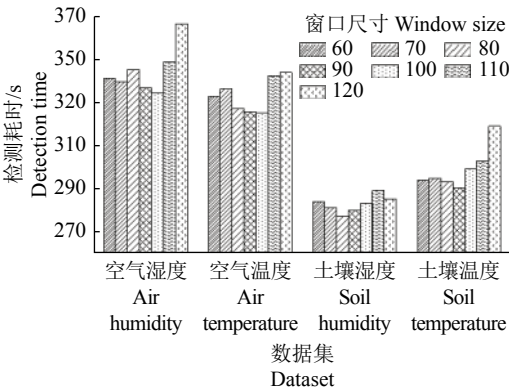


图 6 不同滑动窗口尺寸与数据集的模型检测耗时
Fig. 6 Anomaly detection time of different sliding window sizes and datasets

综上, 采用滑动窗口机制后, 可提升模型对异常数据的检测能力, 针对不同特征数据选取合适的滑动窗口尺寸, 可提升异常检测准确率和降低检测耗时。

2.3 模型异常检测性能

对比分析 SVM、RF、CNN 和 BCNN 模型的数据异常检测性能, 各种模型的异常检测准确率和 F1 值如图 7 所示。BCNN 模型在空气湿度、空气温度、土壤湿度和土壤温度数据集上均表现出良好的异常检测性能, 准确率均在 97% 以上, 其中空气湿度数据集的准确率最高, 达到 99.29%, 优于 SVM、RF 和 CNN 的 95.68%、94.67% 和 93.31%。相较于 CNN 模型, BCNN 模型在空气湿度、空气温度、土壤湿度和土壤温度数据集上准确率分别提升了 5.98%、1.48%、2.19% 和 2.55%; 相较于 SVM 和 RF 模型, BCNN 模型在空气湿度数据集上的准确率与二者差异最为明显, 分别高出 3.61% 和 4.62%。由图 1 可知, 空气湿度数据波动性较大, 周期性特征弱于其他数据集, 故 BCNN 在处理波动性大的数

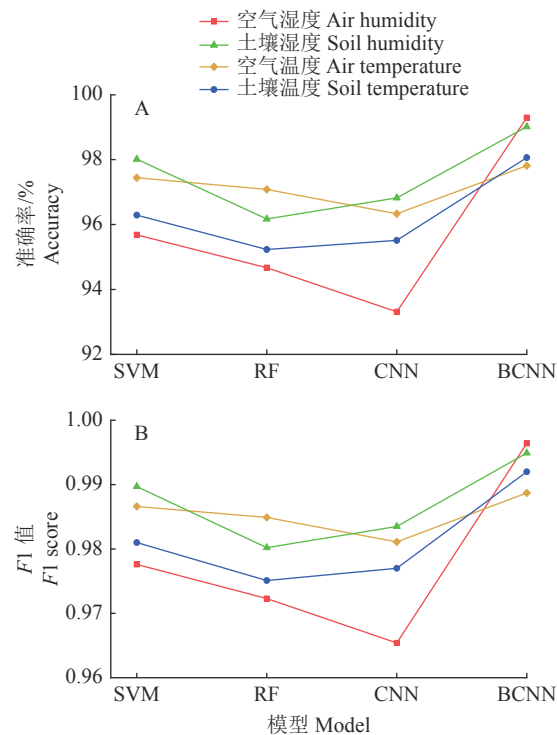


图 7 模型异常检测准确率与 F1 值

Fig. 7 Anomaly detection accuracy and F1 score of different models

据时性能更佳。BCNN 模型在空气湿度、空气温度、土壤湿度和土壤温度数据集上的 $F1$ 值分别为:0.996 4、

0.988 7、0.994 9 和 0.992 0，均优于其他模型。空气温度数据集中，4 种模型的 $F1$ 值差值较小，均小于 0.0076。在波动性较大的空气湿度数据集中，BCNN 模型的 $F1$ 值达到 0.996 4，分别优于 SVM、RF 和 CNN 模型 0.018 8、0.024 1 和 0.031 0。因此，BCNN 模型异常检测性能优于其他模型，且 BCNN 模型对波动性大的数据检测能力更优，而其他模型随着数据波动性增大，异常检测能力有所下降。

总之，相较于传统机器学习算法和深度学习模型，BCNN 模型对异常数据的检测能力在不同特征的数据集上均得到了一定程度的提升。其中，对波动性较大的数据集的异常检测能力最优。

各种异常检测模型检测耗时如表 1 所示。BCNN 模型检测耗时大幅低于同类的深度模型 CNN，仅为 1/6~1/7。相较于 SVM 和 RF 传统机器学习算法，BCNN 模型虽然需要额外的检测时间，但异常检测准确率和 $F1$ 值都有明显提升，增强了模型对波动性较大数据的检测能力，同时，BCNN 模型的超参数数量较少，有效地降低了超参数选择的复杂度。综上，对于农情数据异常检测，BCNN 模型在检测能力和耗时方面优于同类深度学习模型，在检测能力和超参数选择方面优于传统机器学习算法，具有良好的适用性。

表 1 不同模型异常检测耗时
Table 1 Anomaly detection time of different models

模型 Model	空气湿度 Air humidity	空气温度 Air temperature	土壤湿度 Soil humidity	土壤温度 Soil temperature
SVM	10.78	14.62	11.58	14.47
RF	12.02	14.39	12.85	13.77
CNN	2 073.04	2 014.88	1 928.76	2 116.24
BCNN	335.74	329.19	283.32	291.96

3 结论

本研究提出了基于宽度卷积神经网络的农情数据异常检测方法，采用养殖场传感器环境监测数据，综合评估了数据异常检测性能。试验结果表明，基于数据特征的滑动窗口划分子集，可有效地增加模型检测准确率和降低检测耗时。在异常检测模型对比试验中，BCNN 模型取得了最高准确率和 $F1$ 值，对空气和土壤温湿度数据异常检测的平均准确率和 $F1$ 值分别达到了 98.54% 和 0.993 0，相比于 SVM、RF 和 CNN 模型的平均准确率 (96.86%、95.79% 和 95.49%) 和 $F1$ 值 (0.983 7、0.978 1 和 0.976 8)，表现出良好的检测效果。BCNN 模型对波动性较大数据

集的检测能力更强，准确率和 $F1$ 值平均优于其他模型 4.74% 和 0.024 6。另一方面，BCNN 模型检测时间仅为同类深度学习模型 CNN 的 1/6~1/7，且比传统机器学习模型 SVM 和 RF 采用更少的超参数。本研究为农业物联网数据高质量感知及农业信息化提供了一定的参考。考虑到无线传感器网络规模不断扩展，后续研究将更多地关注传感器网络群体异常数据检测，进一步提高农业物联网感知数据质量。

参考文献:

[1] BOURSISANIS A D, PAPADOPOULOU M S, DIAMANTOULAKIS P, et al. Internet of things (IoT) and agricultural unmanned aerial vehicles (UAVs) in

smart farming: A comprehensive review[J/OL]. Internet of Things, 2020. [2021-03-25]. <https://doi.org/10.1016/j.iot.2020.100187>.

[2] LI C, NIU B. Design of smart agriculture based on big data and Internet of things[J/OL]. International Journal of Distributed Sensor Networks. 2020, 16(5). [2021-03-25]. <https://doi.org/10.1177/1550147720917065>.

[3] LIOUTAS E D, CHARATSARI C. Big data in agriculture: Does the new oil lead to sustainability?[J]. *Geoforum*, 2020, 109: 1-3.

[4] CHAMARAJNAGAR R, ASHOK A. Integrity threat identification for distributed IoT in precision agriculture[C]//IEEE. 2019 16th annual IEEE international conference on sensing, communication, and networking (SECON). Boston, MA: IEEE, 2019: 1-9.

[5] 费欢, 肖甫, 李光辉, 等. 基于多模态数据流的无线传感器网络异常检测方法[J]. *计算机学报*, 2017, 40(8): 1829-1842.

[6] ZHAO X, LI P, XIAO K, et al. Sensor drift compensation based on the improved LSTM and SVM multi-class ensemble learning models[J]. *Sensors*, 2019, 19(18): 3844.

[7] WANG Y, YANG A, LI Z, et al. Blind drift calibration of sensor networks using signal space projection and Kalman filter[C]//IEEE. 2015 IEEE tenth international conference on intelligent sensors, sensor networks and information processing (ISSNIP). Singapore: IEEE, 2015: 1-6.

[8] 高鹏, 谢家兴, 孙道宗, 等. 基于物联网和 LSTM 的柑橘园土壤含水量和电导率预测模型[J]. *华南农业大学学报*, 2020, 41(6): 134-144.

[9] 戴泽翰, 郑正, 黄莉舒, 等. 基于深度卷积神经网络的柑橘黄龙病症状识别[J]. *华南农业大学学报*, 2020, 41(4): 111-119.

[10] YOUNG T, HAZARIKA D, PORIA S, et al. Recent trends in deep learning based natural language processing[J]. *IEEE Computational Intelligence Magazine*, 2018, 13(3): 55-75.

[11] WEI J, HE J, CHEN K, et al. Collaborative filtering and deep learning based recommendation system for cold start items[J]. *Expert Systems with Applications*, 2017, 69: 29-39.

[12] HERSHEY S, CHAUDHURI S, ELLIS D P W, et al. CNN architectures for large-scale audio classification [C]//IEEE. 2017 IEEE international conference on acoustics, speech and signal processing (ICASSP). New Orleans, LA: IEEE, 2017: 131-135.

[13] BAO Y, TANG Z, LI H, et al. Computer vision and deep learning-based data anomaly detection method for structural health monitoring[J]. *Structural Health Monitoring*, 2019, 18(2): 401-421.

[14] TANG Z, CHEN Z, BAO Y, et al. Convolutional neural network-based data anomaly detection method using multiple information for structural health monitoring[J]. *Structural Control and Health Monitoring*, 2019, 26(1): e2296.

[15] LIU Z, CHEN C L P, FENG S, et al. Stacked broad learning system: From incremental flattened structure to deep model[J]. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2020, 51(1): 209-222.

[16] WANG Z, OATES T. Imaging time-series to improve classification and imputation[C]//IJCAI. 2015 24th international joint conference on artificial intelligence(IJCAI 2015). Buenos Aires: IJCAI, 2015: 3939-3945.

[17] ZIDI S, MOULAH T, ALAYA B. Fault detection in wireless sensor networks through SVM classifier[J]. *IEEE Sensors Journal*, 2017, 18(1): 340-347.

[18] SHEEBA R, JIJIG, RAJ T A B. WFCM based big sensor data error detection and correction in wireless sensor network[J]. *Cluster Computing*, 2019, 22(2): 3191-3200.

[19] ZHANG T, WANG X, XU X, et al. GCB-Net: Graph convolutional broad network and its application in emotion recognition[J/OL]. *IEEE Transactions on Affective Computing*, Early Access, (2019-08-27) [2021-03-25]. doi: [10.1109/TAFFC.2019.2937768](https://doi.org/10.1109/TAFFC.2019.2937768).

[20] CHERKASSKY V, MA Y. Practical selection of SVM parameters and noise estimation for SVM regression[J]. *Neural Networks*, 2004, 17(1): 113-126.

[21] 段青玲, 肖晓琰, 刘怡然, 等. 基于 SW-SVR 的畜禽养殖物联网异常数据实时检测方法[J]. *农业机械学报*, 2017, 48(8): 159-165.

[22] 杨正理, 史文, 陈海霞, 等. 大数据背景下采用互信息与随机森林算法的空气质量预测[J]. *环境工程*, 2019, 37(3): 180-185.

[23] 熊迎军, 沈明霞, 陆明洲, 等. 温室无线传感器网络系统实时数据融合算法[J]. *农业工程学报*, 2012, 28(23): 160-166.

【责任编辑 李晓卉】