

基于深度学习的大数据异常检测方法研究

单彩云

南京传媒学院传媒技术学院 江苏南京 211172

摘要: 本文针对大数据异常检测关键问题开展系统性研究,采用深度学习技术构建自编码器、循环神经网络及生成对抗网络模型,通过多层非线性映射、隐含空间重构及序列预测实现高维数据特征抽取与异常样本精准判定。模型利用反向传播、梯度下降、正则化约束、批量归一化及Dropout机制优化参数收敛,显著缓解噪声干扰和维数灾难。实验证明三模在降维、密度拟合、时序建模及对抗机制下呈优性能,为风控、网安及监控提供预警支撑力卓越。

关键词: 深度学习; 大数据; 异常检测

引言

随着大数据时代信息复杂性不断提升,异常检测技术已成为网络安全、金融风控和工业监控等领域的核心课题,深度学习凭借其多层非线性映射和高维特征提取能力,为异常识别提供了全新技术路径^[1]。自编码器通过构建隐含空间实现数据自我编码与重构,揭示异常样本在重构误差上的显著差异;循环神经网络借助长短时记忆和门控单元捕捉时序数据内在关联性,精确定位突发性异常模式;生成对抗网络利用生成器和判别器之间的对抗训练,优化数据分布估计,提升异常检测精度。基于深度学习的大数据异常检测方法研究旨在构建系统化理论框架,探索多模型融合优化策略,推动异常检测技术向高效、智能方向发展。

一、深度学习在大数据异常检测中的应用价值

深度学习作为一种多层非线性映射技术,在大数据异常检测领域具备显著的理论优势和应用价值,其核心在于通过构建深层神经网络,实现对高维数据的特征抽取、降维表示与复杂分布建模,从而有效克服传统统计方法在数据维数灾难、噪声干扰与特征冗余等问题。该技术依托卷积神经网络、循环神经网络、自编码器及生成对抗网络等模型架构,充分利用反向传播、梯度下降及正则化机制,实现端到端的自动化特征学习,进而对异常样本进行高精度重构与误差量化。通过构建多层隐

含变量模型,深度学习方法能够捕捉数据内在结构及时序依赖性,利用重构误差、概率密度估计及对抗训练等策略,实现对离群点、异常模式和微弱异常信号的精准判定。同时,采用批量归一化、残差连接、注意力机制及多尺度特征融合等先进技术,有效提升模型鲁棒性和泛化能力,确保大规模数据处理过程中的稳定性与实时性。该方法在大数据环境下通过端对端模型训练,能够实现异常检测的动态阈值设定和系统自适应调控,为金融风控、网络安全、工业监控等领域提供高效智能的技术支持,推动异常检测从静态规则向智能自学习模式转变,体现出深度学习在大数据异常检测中的广泛应用价值和前沿研究潜力^[2]。

二、基于深度学习的大数据异常检测方法研究

1. 自编码器

自编码器作为无监督学习模型的核心架构,通过构建对称型神经网络结构实现输入数据的编码与解码过程,形成高效的低维隐含特征表示,其主要构成包括编码器、瓶颈层和解码器,各层之间采用非线性激活函数和权重共享机制,将高维数据映射至紧凑特征空间,并利用重构误差作为模型性能指标对输入数据进行精细还原,异常样本在重构过程中往往呈现出较大误差,从而实现异常检测功能;为提升鲁棒性和泛化能力,自编码器模型衍生出堆叠自编码器、变分自编码器与去噪自编码器等多种变体,通过增加网络深度和引入噪声扰动,实现对数据分布内在结构的更高精度刻画,利用反向传播算法和梯度下降优化器精细调整网络参数,多层非线性映射、正则化约束以及稀疏编码策略协同作用于特征提取过程中,有效缓解数据冗余与噪声干扰问题;在大数据异常

作者简介: 单彩云, 1980.12, 性别: 女, 籍贯: 江苏阜宁, 民族: 汉, 学历: 本科, 职称级别: 副研究员, 研究方向: 计算机科学与技术。

检测任务中, 自编码器通过重构误差量化数据与模型内在分布之间的偏离程度, 利用批量归一化、激活函数优化和超参数调优技术保障模型稳定性和收敛速度, 展现出自动特征学习与数据降维的优异性能, 为异常样本与正常样本提供明确区分; 自编码器在高维数据处理领域具备自适应性、计算复杂度低、易于集成分布式计算架构等优势, 为进一步结合多模态数据处理、深层网络结构优化及实时异常检测系统构建提供坚实的理论依据和实践路径, 成为当前大数据异常检测技术中不可或缺的核心算法^[3]。

2. 循环神经网络

循环神经网络通过隐含状态的时间传递机制实现序列数据的动态建模, 利用前馈传播与反向传播算法捕捉长短时依赖性, 其变种长短时记忆网络 (LSTM) 与门控循环单元 (GRU) 引入输入门、遗忘门、输出门等关键机制, 有效解决梯度消失与梯度爆炸问题, 实现对历史信息的精细选择性记忆与更新, 确保时序特征提取的高精度与鲁棒性; 在大数据异常检测中, 循环神经网络依托多层堆叠结构和时序状态传递, 通过最小化均方误差、交叉熵等损失函数, 实现对正常状态下数据分布的拟合与预测, 从而将预测误差作为异常评分, 利用残差分析、概率密度估计等方法精准定位异常样本; 模型训练过程中采用批量归一化、Dropout 正则化以及自适应学习率调整机制, 优化网络参数收敛速度与模型泛化能力, 同时结合超参数搜索策略实现训练过程的高效稳定; 此外, 通过构建双向循环神经网络和多尺度时间窗口分析, 实现前向后向信息融合和时间依赖性增强, 进一步提高对复杂时序模式及突发异常事件的检测能力; 在金融风控、网络安全、工业监控等领域, 该方法能够通过精确的时序特征建模和动态误差反馈, 构建实时监控系统, 自动识别异常模式与偏离分布, 推动异常检测算法从静态规则向智能自学习方向发展, 从而实现大规模数据流中隐含模式的高效捕捉与异常状态的快速定位, 展现出循环神经网络在大数据异常检测应用中的显著优势和前沿研究价值^[4]。

3. 生成对抗网络

循环神经网络通过隐含状态的时间传递机制实现序列数据的动态建模, 利用前馈传播与反向传播算法捕捉长短时依赖性, 其变种长短时记忆网络 (LSTM) 与门控循环单元 (GRU) 引入输入门、遗忘门、输出门等关键机制, 有效解决梯度消失与梯度爆炸问题, 实现对历史

信息的精细选择性记忆与更新, 确保时序特征提取的高精度与鲁棒性; 在大数据异常检测中, 循环神经网络依托多层堆叠结构和时序状态传递, 通过最小化均方误差、交叉熵等损失函数, 实现对正常状态下数据分布的拟合与预测, 从而将预测误差作为异常评分, 利用残差分析、概率密度估计等方法精准定位异常样本; 模型训练过程中采用批量归一化、Dropout 正则化以及自适应学习率调整机制, 优化网络参数收敛速度与模型泛化能力, 同时结合超参数搜索策略实现训练过程的高效稳定; 此外, 通过构建双向循环神经网络和多尺度时间窗口分析, 实现前向后向信息融合和时间依赖性增强, 进一步提高对复杂时序模式及突发异常事件的检测能力; 在金融风控、网络安全、工业监控等领域, 该方法能够通过精确的时序特征建模和动态误差反馈, 构建实时监控系统, 自动识别异常模式与偏离分布, 推动异常检测算法从静态规则向智能自学习方向发展, 从而实现大规模数据流中隐含模式的高效捕捉与异常状态的快速定位, 展现出循环神经网络在大数据异常检测应用中的显著优势和前沿研究价值。

4. 深度信念网络

深度信念网络基于层叠受限玻尔兹曼机构建多层概率生成模型, 通过贪心算法逐层无监督预训练实现数据分布的高精度建模与隐含特征提取, 其核心机制在于利用最大似然估计与对比散度算法对各层权重进行精细调整, 从而构造出具备多级非线性映射能力的深层网络结构, 该结构通过堆叠多个受限玻尔兹曼机实现从低维至高维数据空间的递进式特征抽象与融合, 在预训练阶段构建的先验概率模型为后续监督微调提供坚实基础, 进一步采用反向传播算法与梯度下降优化器在全连接网络中进行参数微调, 实现特征表达与数据重构之间的最优匹配; 在大数据异常检测任务中, 深度信念网络利用隐层激活分布的统计特性, 通过构建多层概率图模型对异常样本与正常样本进行区分, 重构误差及概率密度估计成为判定异常的重要指标, 此外, 通过引入正则化约束、批量归一化技术与自适应学习率调整机制, 有效克服过拟合问题与训练不稳定现象, 确保模型在高维数据空间中具备优异的泛化能力; 模型中各层之间采用非线性激活函数及权重共享策略, 进一步强化了特征提取的鲁棒性, 利用高维空间中隐含分布的细微差异, 实现对复杂异常模式的精准捕捉与动态预测, 进而构建基于深度信念网络的异常检测框架, 为金融风控、网络安全与工业

监控等领域提供了高精度、低延迟的智能预警方案，推动了大数据环境下异常检测方法从传统统计分析向深度概率建模的转型升级，展现出深度信念网络在构建多层次、全局信息融合机制中的卓越优势与前沿研究价值^[5]。

5. 一类分类方法

一类分类方法基于仅有正常样本构建紧凑特征空间，通过最大化数据分布一致性与最小化边界扩散实现异常检测，其基本原理在于构造一个超球面或高维闭合区域以覆盖正常样本分布，利用支持向量数据描述（SVDD）及深度SVDD等模型进行特征映射与区域划分，采用核函数技巧与非线性激活机制在隐含层实现数据降维、特征融合与分布拟合，通过最小化目标函数中样本到中心点距离与正则项的加权求和达到异常阈值设定，模型训练过程中引入反向传播算法与梯度下降优化器，在多层网络结构中逐步逼近最优解，从而实现对异常样本的高效区分；此外，通过集成自动编码器与一类分类器，利用重构误差作为判别依据进一步细化异常评分，实现基于隐空间表示的异常定位，优化多尺度特征融合、批量归一化、Dropout正则化等策略确保模型在大数据环境下的鲁棒性与泛化能力，进而突破传统监督学习对负样本依赖的限制，采用单类数据训练方式充分利用样本内部特征统计分布信息，推动异常检测任务从全局概率估计向局部边界优化转变，实现数据重构、密度估计与距离度量的多重融合，构建具有自适应调控与动态阈值设定能力的异常监控体系，为金融风控、网络安全、工业监控等应用领域提供高精度、低延迟的智能预警方案，展现出一类分类方法在大数据异常检测中的前沿理论价值和应用优势。

6. 图神经网络

图神经网络通过构建节点特征嵌入与邻接矩阵传播机制实现复杂图结构数据的深层语义建模，其基本原理在于利用图卷积、图注意力及消息传递机制对节点、边及子图等多层次结构进行多尺度非线性映射，采用邻域聚合、特征融合与谱域滤波方法在隐含空间中实现数据降维与分布重构，进一步通过反向传播与梯度下降优化器对模型参数进行端到端微调，确保局部拓扑关系与全局结构信息的有效整合；在大数据异常检测任务中，图

神经网络利用节点嵌入向量与图谱信息构建异常评分模型，通过节点度、聚类系数、边权分布等统计量对异常节点或异常子图进行精准判定，利用自适应邻域采样、图池化技术及残差连接策略提升模型鲁棒性和泛化能力，采用谱域滤波、卷积核平滑及多头注意力机制对高维图数据中存在的噪声干扰和稀疏性问题进行有效缓解；此外，通过构建多层堆叠图卷积网络和图自编码器实现节点特征重构与异常模式捕捉，融合局部邻域信息和全局拓扑特征，在金融风控、网络安全、社交网络和物联网数据监控等领域实现异常检测系统的实时预警与动态响应，推动异常检测方法从传统结构分析向图结构深度表示转型升级，展现出图神经网络在高维复杂数据环境中构建多层次、全局信息融合机制的前沿理论优势和应用潜力。

结语

随着信息网络的发展，深度学习模型在金融风控、网络安全、工业监控等领域展现出广泛应用价值，为智能监测、异常预警与自适应调控提供坚实技术支持，同时为多模型融合、算法改进与系统集成探索了新思路，充分证明了基于深度学习的大数据异常检测方法在复杂数据环境下实现高效异常识别与动态阈值设定的理论优势和前沿研究意义。

参考文献

- [1] 王玲玲, 李昕, 邢鲁民, 等. 基于数据融合长短期记忆的大型医疗设备异常检测模型研究[J]. 中国医学装备, 2023, 20(04): 134-138.
- [2] 张靖, 孙文举, 尼文斌, 等. 基于深度学习的风洞天平测力试验数据异常检测方法研究[J]. 实验流体力学, 2022, 36(06): 67-73.
- [3] 李娟, 李海璇, 边玲, 等. 基于跨模态深度度量学习的时序数据异常检测[J]. 计算机仿真, 2022, 39(10): 533-537.
- [4] 刘洪瑞. 基于深度学习的时序数据异常检测方法研究[J]. 黑龙江科学, 2025, 16(04): 91-93.
- [5] 王明泉. 基于深度学习的大数据系统日志异常检测方法研究[D]. 北京交通大学, 2023.