

基于 CNN-LSTM-PSO 的私有云故障检测

曹炳尧, 柏杰, 侯佩儒

(上海大学 特种光纤与光接入网重点实验室, 上海 200444)

摘要: 有效对私有云系统进行故障检测对于保障 IT 系统稳定性及开展可靠性信息活动具有重要的实际意义; 为此从私有云系统的历史趋势数据出发, 将卷积网络 (CNN) 和长短期记忆 (LSTM) 循环神经网络结合, 提出了基于粒子群优化算法 (PSO) 的 CNN-LSTM-PSO 的混合模型, 实现对私有云的故障检测; 采用 X11 算法等技术对数据进行预处理, 使用 CNN 网络提取监控指标时序数据的相关特征信息, 并通过训练 LSTM 网络参数建立 CNN-LSTM 预测模型, 设计了 PSO 算法对预测模型进行参数选优, 减小预测误差, 并以高斯正态分布确定阈值范围, 实现故障的精准检测; 通过和传统单一预测模型以及现有的一些组合预测模型的对比, CNN-LSTM-PSO 模型预测后结果的均方根误差、平均绝对误差和平均百分比误差都低于其余模型; 实验结果验证了模型在预测效果上具备更高的精度和更快的预测速度, 在私有云的故障检测中精确性和实时性都具有良好效果。

关键词: LSTM; 故障检测; X11 分解法; CNN 神经网络; PSO 算法; 高斯正态分布; 超参选优

LSTM Private Cloud Fault Detection Based on CNN-LSTM-PSO

CAO Bingyao, BAI Jie, HOU Peiru

(Key Laboratory of Specialty Fiber Optics and Optical Access Networks, Shanghai University, Shanghai 200444, China)

Abstract: Effective fault detection of private cloud systems is of great practical significance to ensure the stability of IT system and to carry out reliability information activities. To this end, starting from the historical trend data of private cloud systems, a hybrid model of CNN-LSTM-PSO based on particle swarm optimization (PSO) is proposed by combining convolutional network (CNN) with long short-term memory (LSTM) recurrent neural network, so the fault detection of private cloud is realized. The X11 algorithm and other technologies are used to preprocess the data, the CNN network is used to extract the relevant feature information of the time series data for the monitoring indicators, and the CNN-LSTM prediction model is established by training the LSTM network parameters, and the PSO algorithm is designed to optimize the parameters of the prediction model and reduce the prediction error, and the threshold range is determined by the Gaussian normal distribution to achieve accurate detection of the fault. Compared with the traditional single prediction model and some existing combined prediction models, the errors of root mean square, mean absolute and mean percentage in the predicted results of the CNN-LSTM-PSO model are lower than that of other models. The experimental results show that the model has higher accuracy and faster prediction speed in the prediction effect, and has good results in both accuracy and real-time performance in the fault detection of private cloud.

Keywords: LSTM; fault prediction; X11 decomposition method; CNN neural network; PSO algorithm; Gaussian normal distribution; multigrid search

0 引言

私有云的高可靠性是保障业务系统及云上作业正常运行的关键因素, 在基于私有云环境本身高可靠的基础上实现对私有云的故障检测可以及时排除故障因素, 进一步增强系统可靠性。目前在故障检测方面大多直接对系统级别故障数据进行检测, 该方式在故障种类单一时可获得较好的预测结果^[1], 但在私有云这类多因素故障的场景中效果并不理想。同时在正常私有云环境下, 单靠运维人员对监控指标、运行日志等海量数据进行分析, 排查潜在风险, 也存在数据处理困难, 排查准确率低, 流程效率缓慢等缺点。

因此, 针对上述问题, 本文根据私有云监控指标的时间序列数据, 采用 X11 分解算法对监控指标数据进行预处理获取特征数据并作为神经网络输入, 将卷积神经网络 (CNN) 和长短期记忆 (LSTM) 循环神经网络融合进行模型构建, 实现对监控指标时间序列的训练和预测。由于 LSTM 网络克服了 RNN 对时间序列预测时存在的梯度爆炸和梯度消失等问题, 目前 LSTM 模型在时间序列上得到了广泛应用, 并且在私有云监控指标的预测中表现也比较突出。

文献 [2] 验证了 LSTM 模型在云环境下对监控指标时间序列比单一模型在预测上具备精度更高的优势。

收稿日期: 2022-03-31; **修回日期:** 2022-04-18。

基金项目: 国家重点研发计划 (2021YFB2900800); 上海市科委项目 (20511102400, 20ZR1420900); 高等学校学科创新引智计划 (111) (D20031)。

作者简介: 曹炳尧 (1985-), 男, 上海人, 博士, 实验师, 主要从事网络数据处理、虚拟化及监控运维方向的研究。

引用格式: 曹炳尧, 柏杰, 侯佩儒. 基于 CNN-LSTM-PSO 的私有云故障检测[J]. 计算机测量与控制, 2022, 30(8): 76-82, 110.

Sudhakar^[3]结合 RNN 和 LSTM 模型对服务器工作负载进行预测, 并通过对比验证了预测效果其优于 ARIMA 模型, 但相较而言其预测时间更长。Zhang 等人^[4]将萤火虫智能优化算法对 LSTM 预测模型进行优化实现对云主机负载的预测, 提高了峰值值处的预测精度。文献 [5] 提出了 GRU-LSTM 组合模型对云计算资源负载进行预测, 提高了预测精度的同时减少了预测时间。文献 [6] 进一步对 LSTM 模型进行优化, 结合 PF 模型提高了突发负载时预测精度。文献 [7] 结合果蝇优化算法对 LSTM 模型进行改进, 通过对果蝇算法的寻优能力的提高加强了预测的精确度, 但是存在突发状况变化的预测精度低的缺点。

上文实验中, 通过将 LSTM 模型和其他算法结合的方式实现时间序列预测, 这种多模型体系往往在预测精度上高于单一模型, 但是一方面部分模型在 LSTM 超参数的选择往往采用试错法或者只进行简单的优化, 而忽略超参数对模型预测效果的影响, 另一方面在简单优化超参数的模型中往往仅采用单一 LSTM 网络结构进行预测, 预测精度又没有复合模型高。本文在将 LSTM 神经网络模型和 CNN 神经网络结合的基础上, 利用 PSO 算法实现超参数的优化工作, 并采用采用高斯正态分布实现确定预测的波动范围, 设定其上下限阈值, 从而对真实数据进行故障检测, 使其在预测结果及故障检测中产生更好的效果。最后通过和单一模型以及多模型结合的预测方法进行多方面对比, 从而展示、分析本文设计算法的优缺点。

1 相关理论与技术

本节主要介绍卷积神经网络 (CNN)、循环神经网络 (RNN)、长短期记忆神经网络 (LSTM) 以及粒子群优化算法 (PSO)。神经网络是一门重要的机器学习技术, 它通过模拟人脑的神经网络以期实现近似人脑功能的人工智能技术, 而粒子群优化算法是群集智能的一种, 为多主体优化系统体系中的一种算法。

1.1 卷积神经网络

卷积神经网络 (CNN, convolutional neural network) 是一种可以同时学习多种特征的深度神经网络, 目前在已被广泛应用于多个领域, 包括图像、语言识别, 资源预测等^[8]。CNN 的网络结构由输入层、卷积层、池化层、全连接层以及输出层构成, 其中卷积层和池化层通过相互配合可以学习并总结历史数据的抽象特征信息, 其中卷积层对输入数据进行滤波工作, 计算提取出数据的建模特征, 具体计算方式如公式 (1), 从而挖掘出数据特征向量关联关系, 池化层会在获取的特征信息基础上保留卷积后的主要特征的同时降低网络复杂度, 提升模型特征信息提取效果。基于 CNN 算法的优越性已经有很多研究人员将 CNN 成功应用于故障诊断中^[9]。

$$F \otimes \omega = \sum_{k=1}^C \sum_{j=1}^{W_j} \sum_{i=1}^{H_j} (F^k(i, j) \omega^k(i, j)) \quad (1)$$

式中, \otimes 表示卷积计算; F 表示卷积层输入数据; ω 表示卷积核的权重参数; C, W_j, H_j 分别为卷积核的通道数、宽度以

及高度。

一般情况下一维 CNN 网络处理时间序列数据具有更佳效果, 二维 CNN 网络则在图像处理方面效果更好, 本文中数据为时间序列数据, 选择一维 CNN 网络。在经过池化层后将数据输入全连接层, 将多维数据转化为一维。一维 CNN 网络结构如图 1 所示。

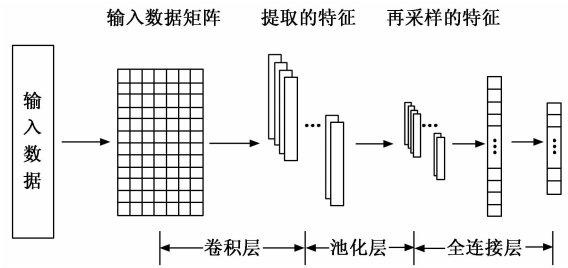


图 1 一维 CNN 结构图

此外, CNN 网络的神经元之间共享一个卷积滤波器的参数, 从而大幅度减少网络冗余参数, 避免过拟合现象。

1.2 循环神经网络

传统的神经网络如 CNN 神经网络在同层之间神经元不会互相传递信息, 输出信号仅与输入信号相关, 因此无法处理具有连续关系的输入信号且无记忆能力, 对本文中的时间序列数据处理效果不佳。为此要构建具备记忆功能的神经网络模型, 可采用循环神经网络 (RNN, recurrent neural network), 通过梯度下降算法、权重共享以及前向反馈机制实现对时间序列的预测, 目前循环神经网络已被广泛应用在自然语言处理^[10]、计算机视觉^[11]、计算机生物学等领域, 在计算机及 IT 系统领域中的应用还较少。

未展开的循环神经网络结构如图 2 (a) 所示, 其网络结构同传统神经网络相同, 包含输入层、隐藏层和输出层。图中 A 表示神经网络模型, x_t 表示输入时间序列在 t 时刻的数据, h_t 表示在 t 时刻的输出数据, 其中区别于 CNN 网络的是隐藏层输出的数据除了会将信息流输出到 h_t , 同时会将当前时刻的神经元信息传递到下一时刻的神经元, 并在下一时刻加以利用, U, W, V 为神经网络中的权重矩阵。这种链式结构揭示了 RNN 本质上是与时间序列相关的, 但是 RNN 虽然理论上可以保留所有的历史时刻的信息, 在实际情况下, 信息的传递会因为时间间隔造成衰减效应, 即梯度消失, 这种情况下会严重降低对数据有长期依赖的信息在处理上的效果。并且 RNN 模型由于在训练时需要预先确定延迟窗口长度, 但是该值的改变难以确定, 其最优值也难以获取。所以本文将进一步研究其优化算法 LSTM。

1.3 LSTM 神经网络原理

由于 RNN 上述的问题, 文献 [12] 提到了长短期记忆神经网络 (LSTM, long short-term memory)。LSTM 通过实现时间递归, 将隐藏层中 RNN 细胞替换为 LSTM 细胞, 并在细胞内采用“门”的概念实现长距离的保留长期历史信息^[13]。图 2 (b) 为 LSTM 神经网络结构图。和 RNN 网络相比, 此处 LSTM 网络添加了表示细胞状态, 用以表示

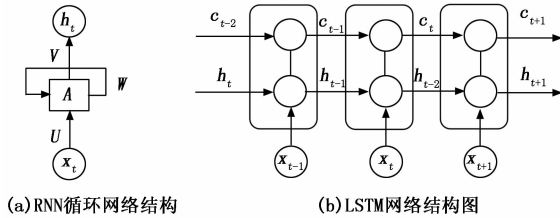


图 2 RNN 和 LSTM 结构图

当前 LSTM 网络的状态信息并用来传递到下一时刻的 LSTM 网络中，其次在隐藏层中设计了 3 个门结构： f_t 遗忘门、 i_t 记忆门和 o_t 输出门，LSTM 神经元隐藏层网络结构如图 3 所示。

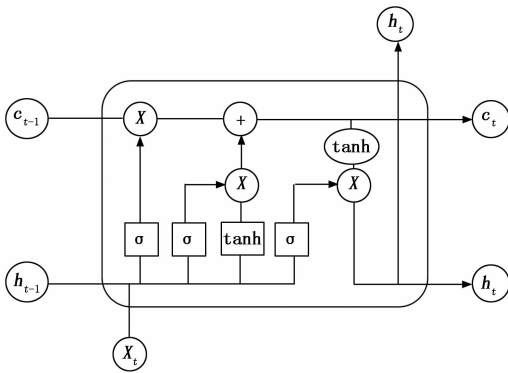


图 3 LSTM 神经元隐藏层网络结构

遗忘门决定了上一周期的时间序列是否从历史数据中进行过滤，记忆门与之相反，它决定输入信息 X_t 和 H_{t-1} 中需要保留的信息，主要将该循环周期内的需要存储的时间序列信息存储到神经元状态中， C_t 细胞状态通过更新传递到下一时刻 C_{t-1} ，输出门并非最终输出状态，仅为当前时刻输出信号，需要作为下一时刻输入信号传递到下一阶段，详细计算方式如公式 (2) ~ (6)：

$$f_t = \sigma(\mathbf{W}_f \cdot [h_{t-1}, x_t] + b_f) \quad (2)$$

$$i_t = \sigma(\mathbf{W}_i \cdot [h_{t-1}, x_t] + b_i) \quad (3)$$

$$c'_t = \tanh(\mathbf{W}_c \cdot [h_{t-1}, x_t] + b_c) \quad (4)$$

$$o_t = \sigma(\mathbf{W}_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t \cdot \tanh(c) \quad (6)$$

式中， \mathbf{W} 是门的权重矩阵， σ 表示 sigmoid 函数， b 是门的偏置项。

由获取到的时间序列数据输入当前时刻数据 X_t ，前一时刻的细胞状态 C_{t-1} 作为该网络神经元状态输入，将当前时刻的信息 X_t 通过遗忘门 f_t 处理，去除冗余信息以后和输入门合并生成当前时刻的细胞状态 C_t ，同时由处理过的历史信息与经过记忆门 i_t 处理后的输入信息进行运算得到当前时刻输出信号 h_t ，该信号也将传递到下一时神经元。

LSTM 神经网络通过增加细胞状态，并经过遗忘门和记忆门更新细胞状态，从而长时间保存了时间序列数据的历史信息，有效缓解了梯度消失、梯度爆炸以及长期依赖的问题。

1.4 粒子群优化算法

在 LSTM 算法中，除了 LSTM 内部网络结构的参数需要通过训练得到，还存在大量需要在训练网络模型之前预设的参数，这些参数被称为超参数 (hyper-parameters)，这些参数在训练过程中保存不变，但其值对模型训练具有较大影响，训练完成后通常需要进行调优，找到一组全局最优超参数。目前超参数优化算法包括网格搜索法、随即搜索算法、贝叶斯算法等，本文采用粒子群优化算法 (PSO, particle swarm optimization) 对 LSTM 模型的超参数进行优化。

PSO 算法是由 Eberhart 和 Kennedy 提出，通过模仿大自然群居动物的集群觅食行为，通过在群体中的个体间的互相关系来寻找最优解^[14]。其凭借易于实现、参数少等优点被广泛应用于神经网络训练、遗传算法等领域。粒子群通过粒子间信息交换保留记忆进行迭代向全局最优解进行演变，获取最优粒子实现寻优。

寻优超参数被包含着粒子群的向量内，设置 M 个参数构成搜索空间创建粒子群，粒子群内包含 N 个粒子，计算具体方式如公式 (7)：

$$\begin{cases} X_i = (x_{i1}, x_{i2}, \dots, x_{iM}), i = 1, 2, \dots, N \\ V_i = (v_{i1}, v_{i2}, \dots, v_{iM}), i = 1, 2, \dots, N \\ P_{best} = (p_{i1}, p_{i2}, \dots, p_{iM}), i = 1, 2, \dots, N \\ g_{best} = (p_{g1}, p_{g2}, \dots, p_{gM}) \end{cases} \quad (7)$$

式中， X_i 表示粒子的位置， V_i 表示粒子的寻优速度， P_{best} 表示第 i 个粒子的搜索到的最优位置， g_{best} 表示迭代后整个粒子群搜索到的最优位置。

本文将 PSO 应用于 LSTM 的超参数优化中，并建立 LSTM-PSO 模型，以寻找超参数的全局最优解。

2 方案设计

根据私有云监控指标时间序列数据的特点，结合第 1 节相关理论，本节提出 CNN-LSTM-PSO 模型的故障检测的方案，首先采用 X11 分解法等技术对时间序列进行数据预处理，并在此基础上选择为 LSTM 模型添加卷积层的方案实现对模型的训练及预测，最后基于多层 PSO 算法完成对超参的选优算法以进一步优化预测效果，并使用高斯正态分布进行阈值选定。

2.1 整体方案设计

本文以 LSTM 神经网络为核心，将实现过程分为七层，包括数据预处理层、卷积层、输入层、隐藏层、网络训练场、超参优化层和输出层，其基本实现框图如图 4 所示。实现过程描述如下：

步骤 1：数据预处理层将短采样周期监控指标时间序列进行平滑处理，并将数据分解得到周期和趋势数据。

步骤 2：卷积层将数据归一化，并使用 CNN 网络模型进行特征学习并提取。

步骤 3：输入层对特征数据进行训练集划分，数据分割以满足 LSTM 神经网络输入要求。

步骤 4: 隐藏层为 LSTM 核心层, 构成网络训练模型核心, 实现网络训练及结果预测。

步骤 5: 网络训练层通过对训练过程中的模型输出和实际数据计算损失模型 loss, 并采用 adam 优化网络算法优化权重矩阵, 完成模型参数的训练。

步骤 6: 超参优化层初始化粒子群, 并使用 PSO 优化算法迭代训练获取全局最优超参数。

步骤 7: 输出层通过训练完成的神经网络对测试数据集进行预测, 同时通过高斯正态分布计算预测值的阈值, 通过比较真实数据是否在该范围内以检测该时刻监控指标数据是否存在异常。

响之外, 还会受到季节、月度变化等因素的影响, 为此本文采用 X11 算法以去除月份和季节因素, 在获得的趋势周期成分中, 在使用经典分解法即公式 (8) 进而获取到精确的周期数据和趋势数据, 其中 X11 算法基本公式如式 (9):

$$Y_t = C_t S_t I_t P_t D_t \quad (9)$$

式中, Y_t 是原始序列, C_t 是趋势周期成分, S_t 是季节成分, I_t 是先验月因子, P_t 是工作日因子, D_t 是不规则成分。

通过结合 X11 分解法和经典分解法获取到精确的周期和趋势数据。X11 分解法 (X11 decomposition) 可以获得完整数据的所有趋势成分, 对于假期、季节等已知的影响因素都有着更好的处理结果。

2.3 基于 CNN-LSTM 网络模型训练及预测

传统的 LSTM 模型虽然可以对中长时间序列进行预测, 但是由于其遗忘门的存在, 在对大规模数据进行训练时, 会遗忘部分重要信息, 造成模型训练效果出现退化现象。而卷积神经网络 CNN 通过其卷积层可以挖掘历史数据中更深层次的特征信息。因此通过 CNN 网络先进行特征信息学习后再传入 LSTM 网络进行时间序列学习将使模型对未来时间数据的数据更精准。

网络训练和预测的核心是隐藏层, 通过训练获取隐藏层矩阵参数, 并以该模型对数据进行预测。经过 2.2 小结预处理后得到处理后的周期和趋势时间序列分别为 $S(t)$ 和 $T(t)$ 。此处以周期序列 $S(t)$ 为例, 设 $S(t) = \{s_1, s_2, \dots, s_n\}$, 在卷积层采用 z-score 对特征值的训练集进行归一化处理, 将其映射为均值为 0, 方差为 1 的标准序列, 以此消除特征单位及尺度等影响, 其计算方式如公式 (10), 完成归一化的数据表示为 $S' = \{s'_1, s'_2, \dots, s'_m\}$ 。

$$z = \frac{x - \mu}{\sigma} \quad (10)$$

式中, x 为实际值, μ 为均值, σ 为标准差, z 为归一化结果。

将归一化后数据 S' 输入卷积层进行特征学习, 通过处理得到适应 LSTM 网络输入格式的数据 $S'' = \{s''_1, s''_2, \dots, s''_m\}$, 然后将其划分为训练集和测试集 $s_n = \{s_1, s_2, \dots, s_m\}$ 和 $S_{sp} = \{s_{m+1}, s_{m+2}, \dots, s_n\}$, 其中 $m < n$, 本文中预测集取 80% 数据, 训练集取 20% 数据。接下来对测试集数据进行数据切割获取到定长的序列, $X = \{x_1, x_2, \dots, x_L\}$, L 为分割窗口大小, 最后将输入到隐藏层即可进行训练。

初始训练模型采用 64 单元的 LSTM 层, 时间步长为 12, 预测长度为 10。同时为防止过拟合的发生, 采用 EarlyStop 和 Dropout 技术, 其中早停法 (EarlyStop) 通过在模型验证集上表现开始下降时停止训练避免过拟合, 随机丢弃层 (Dropout) 通过一定概率的让部分神经元不参加模型训练以减少过拟合现象。损失模型 (loss) 采用均方误差 (MSE, mean-square error) 进行计算, 并以该损失函数最小为目标, 给定初始的超参数, 并使用 Adam 优化算法对隐藏层的网络权重进行更新, 进而获得训练模型并对测试集进行预测。

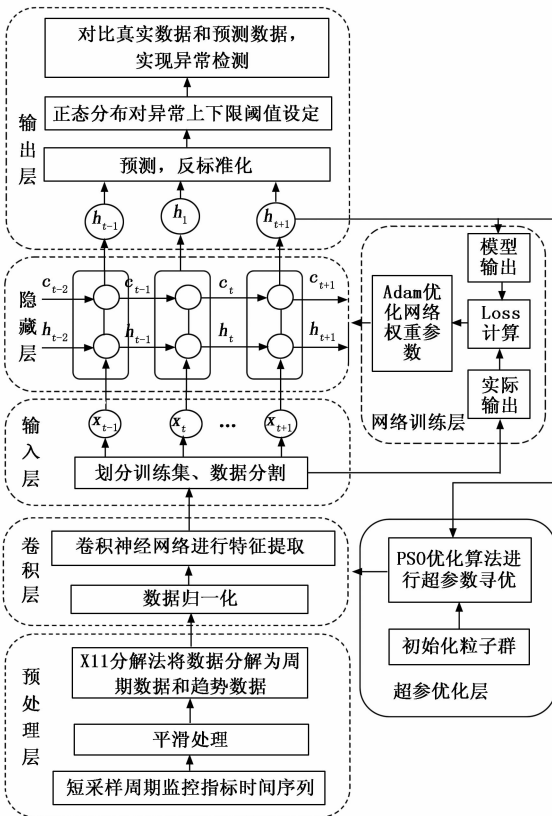


图 4 整体方案设计

2.2 基于 X11 分解法的时间序列预处理

本文监控私有云监控指标其周期粒度一般为 30 s 或者 1 min, 考虑到短周期时间序列的波动性大, 易造成毛刺等现在, 同时为了统一周期粒度指标, 对数据先进行平滑处理, 将周期重置为 3 min, 单周期值取该周期范围内子周期的平均值。同时时间序列通常可以被分解为趋势、周期及残差 3 个变量, 如经典分解法可将时间序列分解为周期成分、趋势成分和残差, 其计算方式如公式 (8):

$$y(t) = S(t) + T(t) + R(t) \quad (8)$$

式中, $S(t)$ 为周期成分 (seasonal component), $T(t)$ 为趋势成分 (trend-cycle component), $R(t)$ 为残差成分 (remainder component)。

但在本文中的监控指标的时间序列除了会被上述因素影

2.4 基于 PSO 算法的超参数优选

LSTM 常见超参数包括学习率 η 、迭代训练次数 C 、分割窗口长度 L 、神经元个数 N 、状态向量大小 S_{state} 等。其中学习率 η 绝对目标函数的收敛能力，合适的学习率可以在合适时间内收敛到局部最小值，反之学习率过大可能会造成模型震荡导致无法收敛，过小则学习效果欠佳，易造成模型训练用时增加，本文学习率的设置范围为 $[0.01, 0.001]$ 。训练次数 C 表示本文需要进行图 3 中网络训练层 Adam 优化网络权重参数的次数，训练次数过小无法得到最优网络权重参数，过多会增加训练时长，浪费计算机计算资源，对实时检测的时效性有一定影响^[15]。

本文将基于 PSO 对 4 个关键超参数（学习率 η 、迭代训练次数 C 、分割窗口长度 L 、神经元个数 N ）进行选优，将这 4 个超参数作为 PSO 粒子，通过更新粒子的速度和位置获取每个粒子全局最优值，LSTM-PSO 模型流程如图 5 所示。

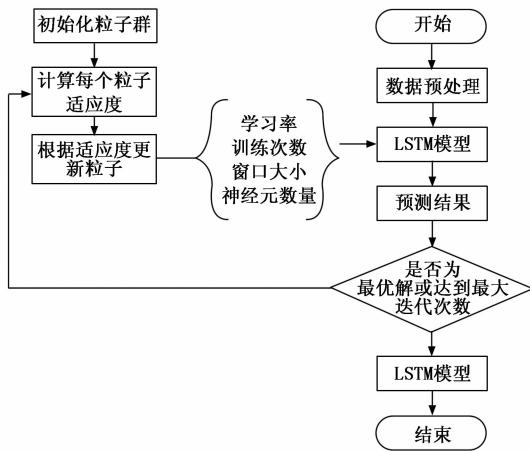


图 5 LSTM-PSO 预测模型流程图

在图 4 的基础上，将 LSTM 模型训练过程置于 PSO 模型内，对于关键的超参数先设置经验值进行模型训练，然后根据表 1 中超参数取值范围通过计算粒子适应度更新粒子迭代训练获取最优解。如表 1 为本文设置超参数取值范围。

表 1 超参数取值范围

参数	最小值	最大值
学习率 η	0.001	0.01
迭代训练次数 C	50	1 000
分割窗口长度 L	1	600
神经元个数 N	1	400

PSO 通过适应度来评价解的好坏，本文中适应度函数通过设置真实值和预测值的平均百分比误差 (MAPE, mean absolute percentage error) 的最小值来实现，公式如 (11) 和 (12)：

$$M = \frac{1}{n} \sum_{i=1}^n \left| \frac{P_{act} - P_{pre}}{P_{act}} \right| \quad (11)$$

$$F(x) = \min(M) \quad (12)$$

式 (11) 中， n 表示预测时间序列的监控项个数， P_{act} 表示监控项的真实值， P_{pre} 表示监控预测值， M 表示 MAPE，式 (12) 中 F 表示适应度。

2.5 基于高斯正态分布的阈值选定

高斯正态分布模型计算简单，且运算速度快，在进行故障实时检测中更快速、准确。通过对样本值的均差和方差来计算整体概率分布，设随机变量服从一个位置参数、尺度参数为的概率分布，且其概率密度函数为式 (13)，则该随机变量为正态随机变量，且符合正态分布^[16]。

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \quad (13)$$

本文进行预测的趋势周期模型符合正态分布。正态分布进行阈值计算求得阈值上限和阈值下限，在本文中必须当趋势周期实际值超过阈值时，将其判定为出现故障。异常检测的实际原理就是小概率事件一般不会发生，当出现时即视为出现故障^[17]。

3 实验测试与分析

3.1 数据集及故障类型

在私有云环境中，按照固定时间采集到监控项指标数据作为时间序列，其中监控项指标包括 CPU、内存等基本监控指标、电源开断、物理服务器风扇转速等物理指标以及应用服务状态、端口状态等监控指标。本文主要以已搭建私有云内某虚拟主机 CPU、内存、磁盘 3 种监控指标作为时间序列进行数据分析和模型对比，每条指标选取 1 年的数据，初始周期统一取 3 min，将前 3 个月数据作为训练集，后 9 个月数据作为测试集。

而在上述指标中呈现的故障类型也存在不同的形式，通常包括点异常、集合异常以及峰值异常，如图 6 所示。其中点异常可以在检测期间单独处理而无需考虑时间关系，所以最好处理，通常简单阈值设定的方法即可处理。峰值异常则取决于其相邻数据点的值，因此需要检测局部信息，而集合异常是由于一系列数据点同时表现出异常而产生的，所以需要时间序列进行处理，本文采用的 CNN-LSTM-PSO 模型对此有着很好的检测效果。

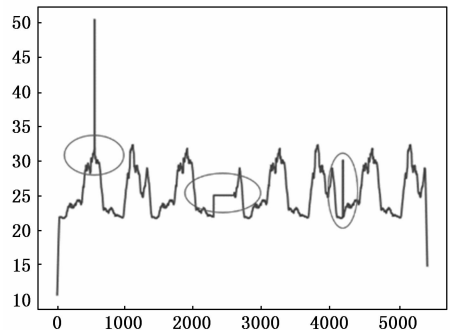


图 6 私有云监控指标故障类型

同时为对模型预测结果进行比较，本文采用标准差 (SD, standard deviation) 来反映实际值与预测值的离散程

度^[18], 模型监控指标的预测值和实际值分别进行均方根误差 (RMSE, rootmeansquareerror)、平均绝对误差 (MAE, mean absolute error) 和平均百分比误差 (MAPE) 的计算对预测结果的精确度进行比较, 这 3 个值越低, 表示误差越小, 精度越高。其中 RMSE、MAE 计算方式如公式 (14)、(15), MAPE 如公式 (11):

$$RMSE = \sqrt{\frac{1}{m} \sum (y_{act} - y_{pre})^2} \quad (14)$$

$$MAE = \frac{1}{m} \sum_{i=1}^m |y_{act} - y_{pre}| \quad (15)$$

式中, y_{act} 表示第 i 个样点监控项的实际值, y_{pre} 表示第 i 个样点监控项的预测值。

故障检测的能力则使用精度 (precision)、召回率 (recall) 和 F1 分数来评估, 精度越高表示检测准确率越高, 召回率越高表示漏检故障可能性越低, F1 越高表示整体效果越好, 计算方式如公式 (16) ~ (18) 所示:

$$Precision = \frac{TP}{TP + FP} \quad (16)$$

$$Recall = \frac{TP}{TP + FN} \quad (17)$$

$$F1 = \frac{2 \times precision \times recall}{precision + recall} \quad (18)$$

式中, TP (truepositive) 表示正确检测出的异常数, FP (false positive) 表示正常数据被检测为故障的数量, FN (false negative) 表示异常数据被检测为正常的数量。

3.2 对比模型

本文将会从以下几个时间序列模型和所提出的 CNN-LSTM-PSO 模型进行实验对比。

1) 自回归移动平均 (ARIMA, autoregressive integrated moving average model) 模型基于时间序列的趋势变化、周期性和随机干扰实现对平稳时间序列的预测^[19]。模型表示为 ARIMA (p, d, q , 其中 p 为自回归项数, d 为差分项数, q 为移动平均项数^[20]。本文基于 ARIMA 模型, 并结合网格搜索法对 ARIMA 模型的 3 个参数进行调优, 从而实现对监控指标的时间序列预测。

2) Holt-Winters 指数平滑不同于常见的时间序列分解法, 其基本思想是通过加权平均移动法来实现时间序列预测^[21]。Holt-Winters 可以在捕捉周期分量和趋势分量的同时直接进一步捕捉到季节分量, 以表示序列数据的局部特征变化, 其原理是通过将时间序列上远近时刻的数据进行权重划分, 其中较近时刻给予较大权值, 较远时刻给予较小的权值从而同时兼顾所有时刻, 实现时间序列的预测^[22]。

3) 将本文第一节提到的 RNN 模型以及单独使用 LSTM 模型的预测进行效果比对。

4) 采用论文 [5] 的模型, 将 GRU 和 LSTM 组合构建 GRU-LSTM 模型进行模型训练和预测。该模型先使用 GRU 网络提取监控指标的输入变量特征, 构成高维映射空间时序性特征向量, 然后将结果输入到 LSTM 网络进行训练。

3.3 平台环境

本文使用计算机配置如下: 处理器为 Intel (R) Core (TM) i7-8750H, CPU 频率为 2.2 GHz 和 2.21 Hz; 内存为 32 GB; 操作系统为 Windows11 (64 位); 程序设计语言 Python3.7, 采用 pandas、keras、states 等程序包。

3.4 实验结果

为直观体现本文模型的预测效果, 以监控指标 CPU 使用率作为预测数据, 选取数据的间隔时间作为横坐标, CPU 使用率作为纵坐标。图 7、图 8 分别为本文设计模型 CNN-LSTM-PSO 和 GRU-LSTM 模型对 CPU 指标进行预测的结果, 可以看到两种模型都可以大致预测到数据的走向, 但本文预测拟合效果更好, 精度更高, 而 GRU-LSTM 在线性上升或者下降部分效果较好, 但在谷峰值位置处明显有部分数据预测效果不佳, 不如本文模型。由此可看到 CNN 结合 LSTM 的模型对极端峰、谷值都有着更好的拟合效果。且在抖动频繁的时间点, CNN-LSTM-PSO 模型的预测结果更接近真实值, 这也证明了通过 PSO 实现超参选择在一定程度上提高了模型的精确度。

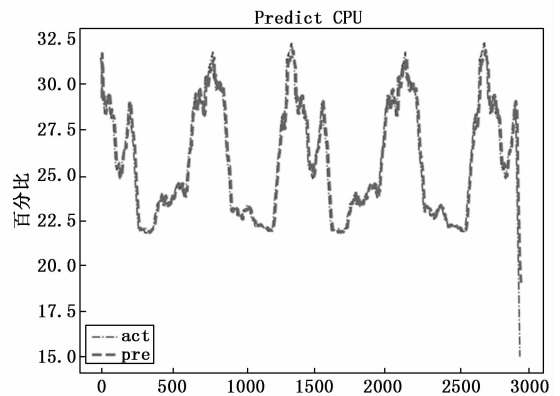


图 7 CNN-LSTM-PSO 模型预测效果

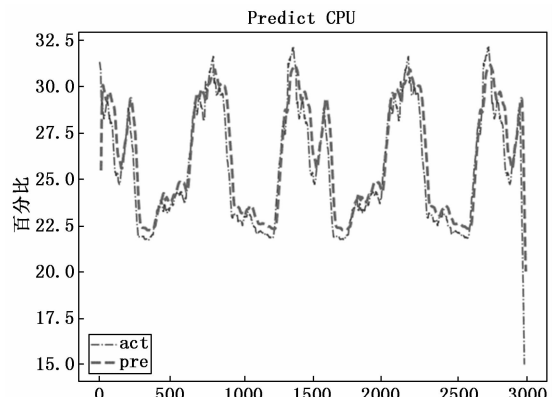


图 8 GRU-LSTM 模型预测效果

为进一步说明 CNN-LSTM-PSO 模型预测性能, 将其分别与 ARIMA、Holt-Winters、RNN、LSTM 以及 GRU-LSTM 五种模型进行对比, 并以 RMSE、MAE 以及 MAPE 作为参考指标, 同时计算模型训练和预测的耗时, 结果如表 2 所示。

表 2 ARIMA、Holt-Winters、RNN、LSTM、GRU-LSTM 及本文模型预测效果对比结果

模型	RMSE	MAE	MAPE/%	耗时/s
ARIMA	2.193	3.14	1.90	12.22
Holt-Winters	1.712	5.20	1.71	1 125.42
RNN	3.189	4.52	4.51	80.52
LSTM	2.231	2.28	2.33	60.25
GRU-LSTM	1.482	2.02	1.77	80.72
本文模型	1.157	1.15	1.21	62.15

从表 2 可以看出, RNN 和 LSTM 神经网络模型误差都比较大, 预测准确度最不理想, Holt-Winters 预测效果略有提升, 但是其耗时过长, ARIMA 模型耗时最短, 但是其预测效果较差, 论文 [5] 中所提的 GRU-LSTM 模型在可控耗时情况下其预测效果也较好, 但是相比于本文的 CNN-LSTM-PSO 模型在预测效果上还有一定差距, 其中对于监控指标波动处的预测效果差距更大。并且本文模型在训练和预测的总耗时上虽比 ARIMA 和 LSTM 模型略高, 但相比于 Holt-Winter、RNN 和 GRU-LSTM 都要略快一些。可以看出, 本文提出的 CNN-LSTM-PSO 在模型性能上有着较好的优化。最后本文通过高斯正态分布进行阈值选定。当实际值超过图中阴影部分上下阈值时, 即判定此时出现故障。

本文通过精度、召回率和 F1 值来比较本文设计模型的故障检测效果, 如表 3 为内存使用率的性能比较结果。由上文预测效果可以分析出单 RNN 模型和 LSTM 模型效果较差, 此处将不在进行比较。

表 3 ARIMA、Holt-Winters、GRU-LSTM 及本文模型故障检测效果对比结果

模型	精确率	召回率	F1 值
ARIMA	0.231	0.4	0.293
Holt-Winters	0.621	0.5	0.553
GRU-LSTM	0.686	1	0.813
本文模型	0.841	1	0.914

结果表明, 相对于本文提出的模型, ARIMA、Holt-Winters 模型的精度、召回率和 F1 值都有所欠缺, 整体效果较差。而 GRU-LSTM 模型和本文模型召回率都较好, 但是本文模型精确率更优, 因此 F1 值也最好。可以看出, 本文提出的 CNN-LSTM-PSO 模型在故障检测效果上有着明显的优势。

3.5 参数选优

在进行超参数选优时, LSTM 网络的超参数范围根据表格 1 中的范围进行设置, 同时设置种子群规模 seed 为 30, 最大迭代次数 40。寻优主要通过 2.4 节中提高的 LSTM 预测模型的参数得结果, 包括适应度 $F(x)$ 和超参数学习率 η 、迭代训练次数 C 、分割窗口长度 L 、神经元个数 N 的变化趋势, 详细结果如如图 9 所示。

通过 PSO 对 LSTM 网络的超参数进行选优, 当迭代次数达到 19 时适应度最优, 且迭代次数最少, 如图 9 (a) 所

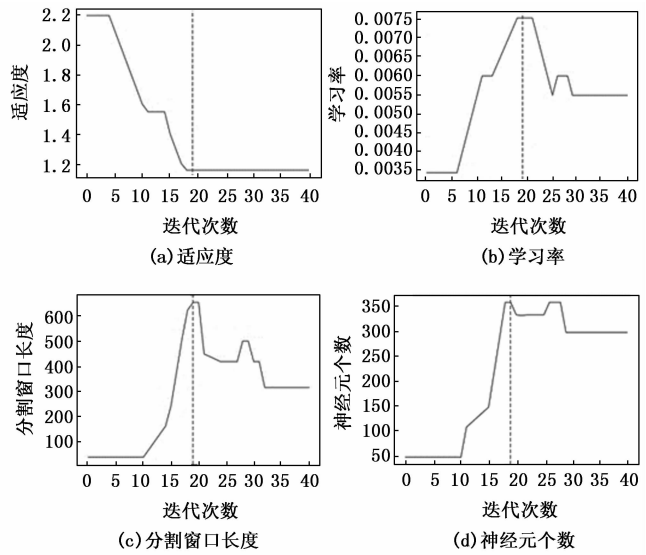


图 9 PSO 优化结果曲线图

示, 此时 3 个超参数都达到最优效果。此时通过图 9 (b)、9 (c)、9 (d) 可以看出学习效率为 0.007 6, 分割窗口长度为 652, 神经元个数为 369。通过对超参数变化曲线图分析可以发现超参数都有两个峰值, 在迭代次数为 19 和迭代次数为 27 分别有一个高峰值, 但是迭代次数在 19 时效果更好。

4 结束语

针对私有云故障检测的问题, 本文提出了 CNN-LSTM-PSO 模型对关键监控指标故障的检测方法, 采用 X11 分解法和经典分解算法可以获取时间序列的周期成分和趋势成分, 同时使用 CNN-LSTM 模型可以对长时间序列预测有更好的拟合效果, 利用 PSO 对 LSTM 模型网络的超参数进行选优工作, 在避免简单人为确定超参数导致无法取得全局最优的同时, 将超参选优的时间控制在确定时间内, 保障模型预测的实时性。采用高斯正态分布确定阈值大小, 最终实现了私有云的故障检测工作。

测试结果表明, 相比较其他时间训练模型, 本文提出的 CNN-LSTM-PSO 模型能够有效提高预测的精确度和准确性, 同时训练及预测耗时在可控范围之内, 在对私有云的故障检测中有着较好效果。

参考文献:

[1] 王鑫, 吴际, 刘超, 等. 基于 LSTM 循环神经网络的故障时间序列预测 [J]. 北京航空航天大学学报, 2018, 44 (4): 772-784.

[2] 崔峰, 郭刚, 饶伟, 等. 基于 LSTM 的云环境异常智能检测方法研究 [J]. 铁路计算机应用, 2020, 29 (6): 44-48, 53.

[3] SUDHAKAR C, KUMAR A R, SIDDARTHA N, et al. Workload prediction using arima statistical model and long short-term memory recurrent neural networks [C] //2018 International Conference on Computing, Power and Communication Technologies (GUCON). IEEE, 2018: 600-604.

(下转第 110 页)