Computer Science and Application 计算机科学与应用, 2020, 10(10), 1863-1869
Published Online October 2020 in Hans. http://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2020.1010196

Hans汉斯

基于PSO-LSTM的网络安全态势感知预测方法

袁永晖

同济大学，上海
Email: 836171392@qq.com

收稿日期：2020年10月7日；录用日期：2020年10月22日；发布日期：2020年10月29日

摘　要

针对传统网络安全态势感知预测方法预测精度低的问题，本文提出了一种基于PSO-LSTM的网络安全态势预测方法。网络安全态势数据由于其具有时序性，过往的态势值往往影响到未来的网络状况。LSTM网络可以有效地处理具有时序性特征的数据，同时依靠PSO算法优秀的全局搜索能力优化LSTM网络的超参数。仿真实验表明，本文提出的方法有效地提高了态势值预测的准确度。

关键词

态势评估，态势预测，粒子群优化，长短时记忆网络

Research on Network Security Situation Prediction Algorithm Based on PSO-LSTM Neural Network

Yonghui Yuan

Tongji University, Shanghai
Email: 836171392@qq.com

Received: Oct. 7th, 2020; accepted: Oct. 22nd, 2020; published: Oct. 29th, 2020

Abstract

Aiming at the low accuracy of the traditional network security situation perception prediction method, this paper presents a network security situation prediction method based on PSO-LSTM. Due to the timing of network security situation data, the past situation values often affect the future network situation. LSTM network can effectively process data with temporal characteristics, and at the same time, the super parameters of LSTM network can be optimized by the excellent

文章引用: 袁永晖. 基于 PSO-LSTM 的网络安全态势感知预测方法[J]. 计算机科学与应用, 2020, 10(10): 1863-1869.
DOI: 10.12677/csa.2020.1010196

global search capability of PSO algorithm. Results show that the method proposed in this paper can effectively improve the accuracy of situation value prediction.

Keywords

Situation Assessment, Situation Prediction, Particle Swarm Optimization, Long and Short Time Memory Network

1. 引言

近年来，随着互联网的日益进步，网络攻击手段也变得越来越难以捉摸。传统的被动防御模式已经难以满足基本的网络防御要求。网络安全态势感知作为一种能够基于环境的、动态、整体地洞悉网络安全风险的能力，从全局视角提升对安全威胁的发现识别、理解分析、响应处置能力的一种方式，成为了当下的研究热点。网络安全态势预测是态势感知最重要的一环，可以实现对未来网络运行状况的预测。传统网络安全态势预测一般基于时间序列和集成技术等进行预测。随着深度学习的不断发展，越来越多的学者将神经网络应用到态势预测的问题中来。文献[1]提出了基于 GA_SVM 的网络安全态势预测模型，该算法首先将重构后的网络安全态势数据输入基于 SVM 网络安全态势预测模型进行训练，然后采用遗传算法获取较优的预测模型参数，解决了传统支持向量机网络安全态势预测模型存在参数选取困难的问题。文献[2]利用 PSO 算法改进了传统 RBF 神经网络，改进了 PSO 的惯性权重，在保证全局寻优的同时又增强了局部搜索能力。文献[3]提出了基于灰色理论和 BP 神经的网络安全态势预测方法。依据灰色模型系数的取值大小选择最合适的背景值，并构造了一种新的模型背景值函数。结合 GM 模型预测网络安全态势，并用 BP 神经网络对态势预测值进行修正。文献[4]提出了基于深度自编码器的网络安全态势预测。通过深度自编码对特征的提取和恢复能力，有效学习到了安全态势数据的稀疏表示，从而提高了安全态势值的预测精度。文献[5]提出了基于人工免疫算法的网络安全态势预测，采用了人工免疫算法来优化网络的结构和神经网络的态势预测值，进一步提高了模型的检测率。

由于网络安全态势数据的时序性，本文提出了一种基于 PSO-LSTM 的网络安全态势预测方法。LSTM 可以有效地学习到长序列数据中所包含的信息。而 LSTM 和其他神经网络一样，模型参数往往要靠人工选择，这就导致模型难以训练。故采用 PSO 算法优化 LSTM 网络，合理确定 LSTM 的模型参数。实验证明，PSO 有效提高了 LSTM 网络的收敛速度和预测精度。

2. 基于 PSO-LSTM 的网络安全态势感知预测方法

2.1. LSTM 算法

循环神经网络(RNN)是一种专门用来处理序列数据的神经网络。和一般神经网络不同的是，它能有效的提取到序列之间的信息。比如单词会因为上下文的不同而导致词义发生变化。但传统循环神经网络面对长序列时容易发生梯度消失以及梯度爆炸的问题。为了解决 RNN 所存在的问题，LSTM 作为它的一种变种被提了出来。

长短时记忆网络(LSTM)主要用于解决 RNN 中梯度消失和爆炸的问题，因此在面对长序列数据时有着良好的表现。其网络结构如图 1 所示。

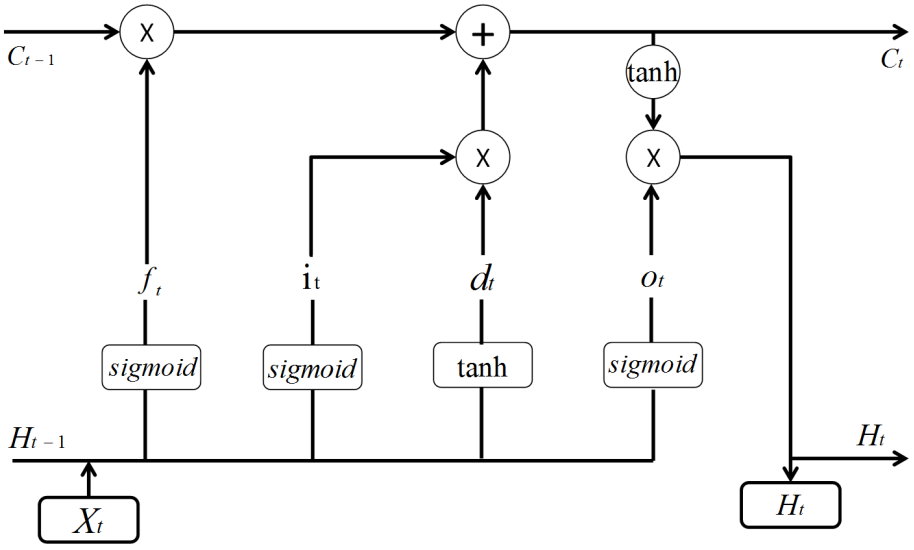

Figure 1. LSTM structure
图 1. LSTM 结构图

相较于传统 RNN 只会传递 C_t，LSTM 同时还会给下一个时刻传递当前的细胞状态 H_t。也就是说比起 RNN，LSTM 不仅会告诉下一个时刻当前的细胞状态，同时还会传递隐层的状态，这为后续一系列的门运算提供了基础。

LSTM 主要包含了三个门：遗忘门、输入们、输出门。

1 遗忘门：这是 LSTM 的遗忘阶段，主要时为了忘掉长序列数据中不重要的信息。由于激活函数是 sigmoid 所以输出在 0 和 1 之间。其中 1 表示保留，0 表示遗弃。其更新公式如公式 1 所示：

$$f = \sigma\left(W_f \cdot \left[H_{t-1}, X_t\right] + b_f\right) \tag{1}$$

2 输入门：这是 LSTM 的选择记忆阶段，这个阶段会将输入有选择性地进行“记忆”。主要是会对输入 X_t 进行选择记忆。重要的信息多记一些，不重要的则需要少记。激活函数主要有 sigmoid 和 tanh。公式如 2~4 所示

$$i_t = \sigma\left(W_f \cdot \left[H_{t-1}, X_t\right] + b_f\right) \tag{2}$$

$$\tilde{C} = \tanh\left(W_C \cdot \left[H_{t-1}, X_t\right] + b_C\right) \tag{3}$$

$$C_t = f_t * C_{t-1} + i_t * C_t \tag{4}$$

3 输出门：这是 LSTM 的输出阶段。这个阶段将要决定哪些信息会被当做当前状态而输出。此阶段结合前面两个阶段给出的信息做出判断。此阶段的激活函数也有两个为 sigmoid 和 tanh。其更新公式如 5~6 所示：

$$o_t = \sigma\left(W_o \cdot \left[H_{t-1}, X_t\right] + b_o\right) \tag{5}$$

$$h_t = o_t * \tanh\left(C_t\right) \tag{6}$$

训练 LSTM 网络可以应用时序性倒传递算法，依据错误修改每次的权重。梯度下降法在循环网络神经网络(RNN)会导致误差梯度随着事件间的时间长度成指数般的消失。但是当设置了 LSTM 单元时，误差也随着倒回计算，从输出影响回输入阶段的每一个门，直到这个数值被过滤掉。因此正常的倒传递类神经是一个有效训练 LSTM 区块记住长时间数值的方法。LSTM 的内部结构主要就是通过门控状态来控制传输状态，能记住关键的信息，忘记长序列信息中不需要的信息，不同于普通的 RNN 那样仅仅只有一

种记忆叠加方式。这对网络安全态势预测这种长序列数据来说，是一种非常有效的网络。

2.2. PSO 算法

粒子群优化(PSO, particle swarm optimization)算法时一种基于生物启发的群体智能优化算法，源于对鸟类捕食的研究。本文采用 PSO 算法对 LSTM 中的超参数进行寻优，从而得到整个网络的最优参数组合。

对于一个 j 的优化问题，种群中共有 i 个粒子，对于每个粒子都有 j 维度的速度和位置两个向量。模拟整个鸟群的飞行过程，不断调整第 i 个粒子的最优位置，最终找到整个粒子群最优的位置。应用于 LSTM 中，其具体步骤如下所示：

步骤 1：初始化算法相关参数，确定种群规模，迭代次数，惯性权重和学习因子的变化区间。

步骤 2：随机生成一个三维的种群粒子，每个维度都是我们更新模型的误差参数同时并初始化粒子的位置和速度。所以粒子的维度就是我们需要优化的参数。

步骤 3：以 LSTM 模型的损失函数作为粒子的适应度函数，适应度函数越大说明模型的损失。

函数越大，所以我们要尽量使适应度变小。这样求得的粒子参数组合越好。计算每个粒子的适应度函数，根据适应度函数确定个体最佳位置和群体最佳位置。

步骤 4：迭代过程中不断更新粒子的位置和速度，再重新算出每个粒子的适应度函数并更新个体最佳位置和群体最佳位置。

步骤 5：当达到迭代次数或粒子的适应度函数趋于稳定时则停止，作为群体最佳位置的粒子为本次求得的最优参数组合，否则转步骤 4 继续迭代。其流程图如图 2 所示。

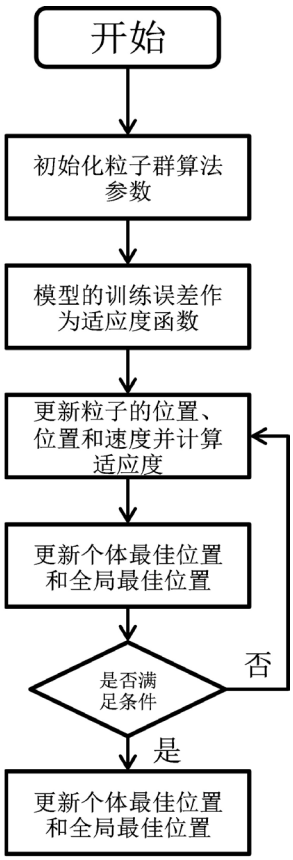

Figure 2. The flow chart of PSO
图 2. PSO 流程图

3. 仿真实验及结果分析

3.1. 实验环境和数据集介绍

实验编程语言使用 Python，版本为 3.7.5。基于 tensorflow、sklearn 等开源库实现。操作系统使用的是 Windows 10、CPU 2.9 GHz、内存 6.0 GB 的 PC 机。数据集选用 KDD-CUP1999 数据集。此数据集是 KDD 竞赛在 1999 年所示用的数据。该数据集模拟了美国空军局域网上采集来的 9 个星期的网络连接数据，分成具有标识的训练数据和未加标识的测试数据。测试数据和训练数据有着不同的概率分布，测试数据包含了一些未出现在训练数据中的攻击类型，这使得入侵检测更具有现实性。

数据集中主要包含 5 种类别的数据包括正常数据和另外四种攻击数据。四种攻击分别为：Probe、denial of service (DoS)，user to root (U2R)，and remote to local (R2L)。四类攻击类型又可详细划分为 39 类具体攻击。NSL-KDD 每条样本包含了 41 个特征属性标签，分别由 38 个 numeric 特征和 3 个 nominal 特征组成。数据集本身极度不平衡，部分攻击类的样本远远少于正常数据。

3.2. 数据预处理

数据集中包含了连续性和离散型的数据。对于连续性的数据为了避免其量纲的不一致性，需要对其进行归一化处理。计算方法如公式 7 所示：

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{7}$$

而对于离散型的数据需要对其进行 one-hot 编码处理。在 41 个固定的特征属性中，9 个特征属性为离散型，对这 9 个特征进行编码后共得到 122 维的特征。同时我们采用文献[6]中的态势评估方法得到了网络安全态势值，并对其进行归一化，部分态势值如图 3 所示。

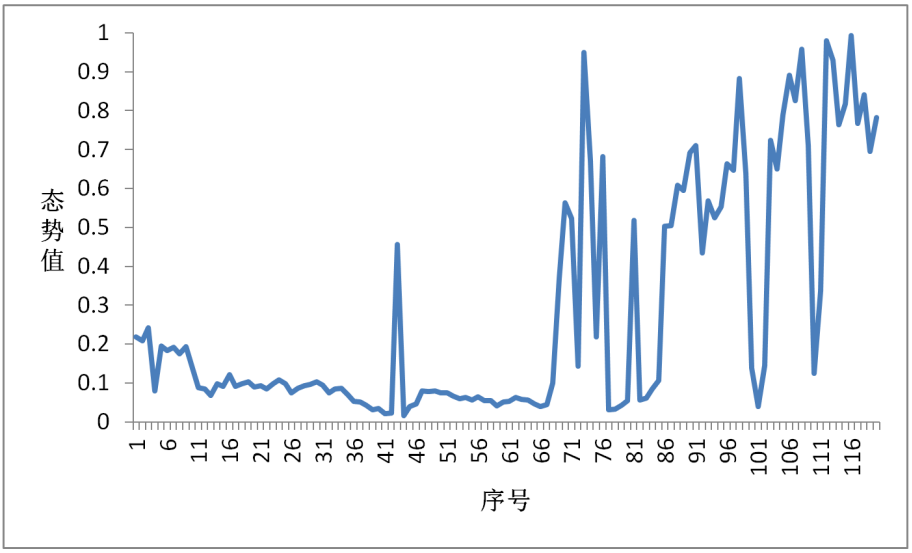

Figure 3. Situation curve
图 3. 态势曲线图

3.3. 评价标准

本文选取均方误差(mean-square error, MSE)对网络安全态势值的预测进行评价。它是衡量真实值与预测值比较常用的一个评价标准。MSE 越大，表明算法的准确度越小，相反，MSE 越小则表明预测值更加

精准。其公式如式 8 所示：

$$\mathrm{MSE}=\frac{1}{n}\sum_{t=1}^{n}\left(x_t-x_t^{\star}\right)^2 \tag{8}$$

其中，x_t 为真实值，$x_t^{\star}$ 为预测值。

3.4. 结果比对和分析

本文分别与 BP、PSO-SVM 进行对比。取测试集中编号 1~10 的数据，分别用三种算法得出了预测值，结果如图 4 所示。

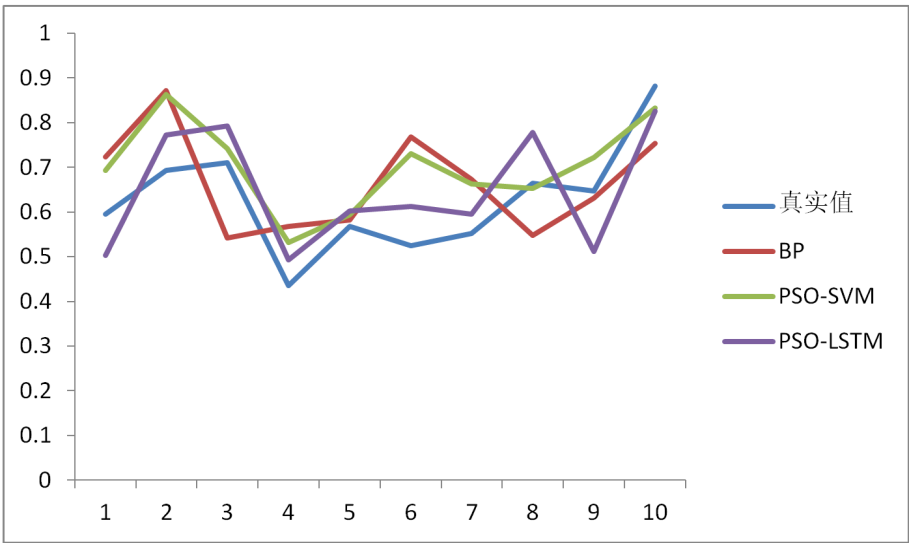

Figure 4. Experimental result diagram
图 4. 实验结果图

从图 4 可以看出，本文所提出的方法与真实数据有着最高的拟合度，数据变化趋势基本一致。BP 神经网络的预测效果最差，其无法捕捉到态势数据所存在的序列信息。可见 LSTM 相比于传统神经网络在态势预测上有较大优势。PSO-SVM 显示出了和 BP 相差不远的精度，由于本文所使用的数据集样本数量巨大，SVM 显然无法有效处理。最终，三个算法的均方误差由表 1 给出。

Table 1. Mean square error table
表 1. 均方误差对照表

算法	BP	PSO-SVM	PSO-LSTM
均方误差	0.0199120	0.0112750	0.006988

4. 结论

本文针对传统网络安全态势预测技术预测精度不高的问题，提出了一种基于 PSO-LSTM 的网络安全态势预测模型。实验表明，相较于 BP 以及 PSO-SVM 算法，本文所提出的算法得到了较高精确率。由于本文所使用的数据集不平衡的特性，下一步将探究如何进一步提高攻击不平衡的样本数据的预测精度。

参考文献

[1] 郭政, 张锐. 基于 GA_SVM 的网络安全态势预测模型研究[J]. 信息通信, 2018(7): 64-66.

[2] 江洋, 李成海, 魏晓辉, 李志鹏. 改进 PSO 优化 RBF 的网络安全态势预测研究[J]. 测控技术, 2018, 37(5): 56-60.
[3] 邓勇杰, 文志诚, 姜旭炜. 基于灰色理论和 BP 神经的网络安全态势预测[J]. 微型机与应用, 2015, 34(20): 1-3+8.
[4] 张生顺. 基于深度自编码器的网络安全态势预测[J]. 信息安全研究, 2019, 5(7): 644-648.
[5] 秦丽娜. 基于人工免疫算法的网络安全态势预测[J]. 长春工程学院学报(自然科学版), 2018, 19(1): 103-106.
[6] 罗昭. 基于神经网络的网络安全态势评估与预测技术研究[D]: [硕士学位论文]. 西安: 西北大学, 2018.