基于 NAWL-ILSTM 的网络安全态势预测方法

朱 江 陈 森

(重庆邮电大学通信与信息工程学院移动通信技术重庆市重点实验室 重庆 400065)

摘 要 安全态势是网络安全预警的前提。各种复杂网络环境中的网络攻击行为给网络带来了意想不到的挑战,导致网络负载增加和网络故障等突发网络安全事件随时都会发生。因此,针对网络安全态势时间序列的不确定性、非线性等特点,为了提高网络安全态势预测的精度,提出了基于改进 Nadam 和改进长短期记忆网络(NAWL-ILSTM)的网络安全态势预测方法。首先,利用一种在线更新机制改进长短期记忆网络(LSTM)以建立态势时间序列预测模型,它可以实时地对接收到的在线观测数据进行参数更新,使代价函数最小化,从而解决了传统 LSTM 网络模型不能合理地利用网络系统在线传送数据的问题,在优化参数更新的同时也大大提高了 LSTM 模型的预测精度;然后,针对神经网络训练过程中收敛速度较慢和训练成本较高的问题,采用 Look-ahead 方法对 Nesterov 加速梯度的自适应估计动量算法(Nadam)的更新公式进行改进,以加快模型的收敛速度,从而加快了 ILSTM 预测模型的训练速度,减少了训练的时间和成本。基于 Python 在 tensorflow 环境下进行仿真实验,结果验证了所提的基于在线更新机制的 LSTM 预测模型的合理性,通过收敛性分析和算法对比得出了 NAWL 算法具有更快的收敛速度的结论。最后,与其他预测模型的对比结果表明了 NAWL-ILSTM 预测模型在态势时间序列分析中具有更强的适用性和更高的准确性。

关键词 网络安全态势预测,长短期记忆网络,在线更新参数,前瞻性技术,适应性动量算法

中图法分类号 TP393 **文献标识码** A **DOI** 10.11896/jsjkx.180901820

Network Security Situation Prediction Method Based on NAWL-ILSTM

ZHU Jiang CHEN Sen

(Chongqing Key Lab of Mobile Communications Technology, School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

Abstract Security situation is the premise of network security warning. The network attacks in complex network environment bring unexpected challenges, causing the sudden network security incidents such as increasing network load and network failure happen at any time. Therefore, taking into account the uncertainty and non-linearity of network security situation time series, in order to further improve the forecast accuracy of network security situation, this paper proposed a network security situation prediction method based on NAWL-ILSTM (Nadam with Look-ahead and Improved Long Short-Term Memory). Firstly, an online updating mechanism is adopted to improve the LSTM to establish time series forecasting model, which can conduct parameter updating in real time for the received online observed data and minimize the cost function, thus solving the problem that traditional LSTM algorithm can't use network system to transmit data online reasonably, further, optimizing the parameter updating and improving the forecast accuracy of LSTM model. Then, aiming at the problems of slow convergence speed and high training cost in the training process of neural networks, the Look-ahead technology is used to improve the updating formula of Nesterov acceleration gradient adaptive estimated momentum algorithm (Nadam) to accelerate the convergence speed of the model, and then the training speed of ILSTM prediction model can be accelerated to reduce training time and cost. The simulation experiments based on Pythonin tensorflow environment demonstrate the rationality of the LSTM prediction model based on online updating mechanism. Convergence analysis and comparison experiments show the NAWL algorithm has faster convergence speed. Finally, the comparison experiments show that the proposed model based on NAWL-ILSTM has stronger applicability and higher applicability in situation time series analysis compared with other prediction model.

到稿日期:2018-09-28 返修日期:2019-02-13 本文受国家自然科学基金资助项目(61271260,61301122),重庆市科委自然科学基金项目(cstc2015jcyjA40050)资助。

朱 江(1977—),男,博士,副教授,主要研究方向为通信理论与技术、信息安全技术等;陈 森(1994—),男,硕士生,主要研究方向为网络安全态势感知,E-mail:1198534370@qq.com。

Keywords Network security situation prediction, Long short-term memory, Online observation data, Look-ahead technology, Adaptive momentum estimated algorithm

1 引言

随着互联网技术的快速发展,计算机网络已经成为了一种不可或缺的通信手段。但是,网络环境中存在着各种各样的威胁,虽然目前已经开发了防火墙、入侵检测系统、病毒杀伤等技术,但这些方法只能应对所发生的威胁,不能很好地控制网络的整体趋势。在此背景下,针对网络安全问题,研究者提出了网络安全态势感知。网络安全态势是网络安全状况的一种趋势。根据网络环境的变换,网络管理员可以采取措施来避免网络攻击或减少网络攻击造成的伤害。网络安全态势预测是一种主动防御机制[1],其首先对现在以及以往的网络态势要素进行分析和理解,然后对将来的网络态势进行推测。由于网络安全的现状是由态势评估之后得到的态势值所反映的,态势值表示每一时刻网络的状态值,因此态势预测问题实际上是时间序列预测问题[2]。由于网络安全态势变化的趋势具有非线性、时变性等特征,因此很多经典的时间序列预测方法很难准确地找出网络的现状与发展趋势之前的关系,导致预测精度无法被提高。

随着人工智能算法的普遍应用,很多研究者提出了预测方法,如马尔可夫链(Markov)[3]、支持向量机(SVM)、神经网络(Neural Network)等。但是,许多研究通过实验发现上述方法都存在不足之处:文献[4-5]中的 RBF 利用优化算法优化深度神经网络模型,虽然很大程度上解决了网络参数的最优问题,但是预测的精度不足,收敛速度过慢;文献[6-7]将 SVM 作为预测方法,由于网络规模大、训练样本复杂,SVM 很难对大规模训练样本进行数据处理,收敛速度慢;文献[8-9]利用 Adam 优化传统的 LSTM 模型来对时间序列进行预测,其没有充分地利用网络实时接收的数据,并且存在预测精度不足、泛化能力弱等问题。

为了解决传统预测模型存在的问题,文中提出了一种基于 NAWL-ILSTM 在线更新网络参数机制的安全态势预测模型。考虑到网上实时在线传送的数据会导致时间序列的增加,本文以历史数据为基础建立模型参数并对其进行优化,将传统的 LSTM 模型改进为利用在线数据更新网络参数机制的ILSMT 模型;同时,本文利用 Look-ahead 方法来改进 Nadam 优化算法,不仅提高了收敛速度,而且大大减少了网络训练的成本。通过将典型的预测模型与 NAWL-ILSTM 预测模型进行比较,验证了本文模型在网络实际情况下具有更高的预测精度以及更快的收敛速度。

2 网络安全态势的系统模型

在网络安全态势预测过程中,存在对网络安全运行产生影响的外界因素,导致网络环境不断复杂化。为了从各个方面检测网络安全的发展走势,本文提出了一种精确的预测和评估网络安全态势的系统模型,如图 1 所示。首先,利用文献[10-11]对某一时间段的安全状况进行定量分析;然后,针对相关的因素进行权重分析,评估出当前状态,并形成网络态势时间序列图;最后,根据历史信息和当前信息预测网络安全的发展趋势。通过 Nadam-ILSTM 网络处理复杂的网络安全状况,利用在线数据传送和改进的 Adam 优化算法来提高模型的优化速度和泛化能力,以更高的精度和更快的速度来预测网络态势时间序列。

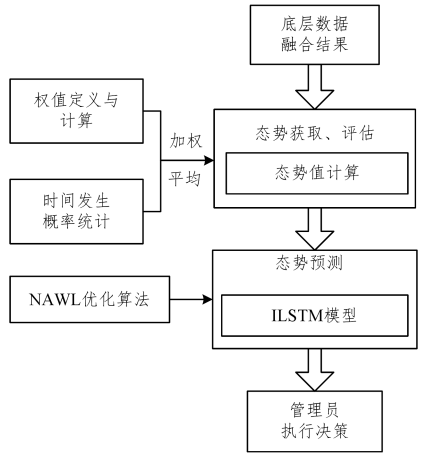

图 1 网络安全态势感知模型

Fig. 1 Network security situation awareness model

3 基于 ILSTM 的态势预测模型

循环神经网络(Recurrent Neural Network, RNN)[12]是一种改进的多层神经网络,由输入层、隐藏层、输出层组成,可以有效解决长期依赖的问题。RNN 可以被认为是同一神经网络的多次复制,每个神经网络单元将消息向下传播。从图 2 可以看出 RNN 是一个链式结构,链式结构特征表明它与时间序列相关。当前时刻的输出不仅与输入时间有关,还与之前的输出有关,这使得之前的信息可以向后传播,这就是 RNN 能够解决时间自相关问题的原因。然而,RNN 最明显的问题是训练过程中的梯度需要反向传播,当输入的时间序列较长时,会出现梯度消失和梯度爆炸的问题。

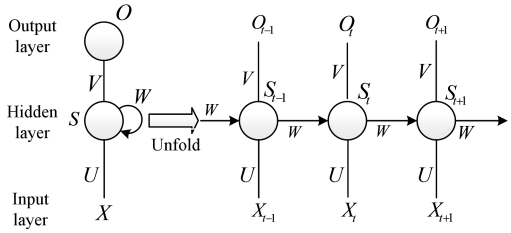

图 2 RNN 隐藏单元的展开图

Fig. 2 Expansion diagram of RNN hidden layer

长短期记忆网络(Long Short-Term Memory, LSTM)[13]可以通过引入内存单元来解决上述 RNN 的问题,使网络学会适时地忘记历史信息,并用新的信息更新内存单元。LSTM 和 RNN 有相同的链式结构,标准 RNN 的隐藏层只有 1 个简单的 tanh 层,而 LSTM 有 4 个交互层。LSTM 有 3 个门,分别是遗忘门、输入门、输出门,可以用来控制每个单元的状态。LSTM 隐藏层的细胞结构如图 3 所示。

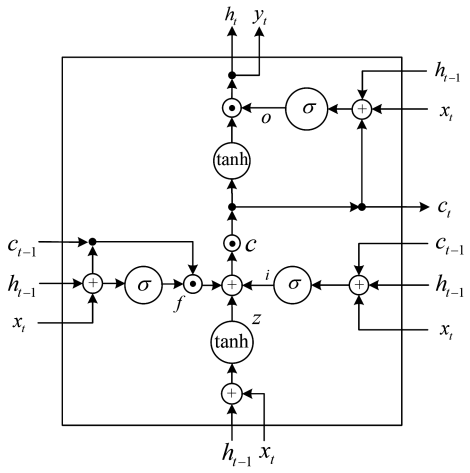

图 3　LSTM 隐藏层的细胞结构

Fig. 3　Cell structure of LSTM hidden layer

然而,LSTM 在网络安全态势预测方面的应用非常少见。现实中,网络一直处于被攻击的状态,通常把从防火墙、IDS 等处收集到的历史信息作为训练数据。因此,为了利用实时接收到的网络态势值和优化网络参数,本文提出了利用改进的 LSTM(Improve LSTM,ILSTM)网络来建立有效的网络态势预测模型。改进的 LSTM 网络是一种最小化代价函数的 LSTM 学习在线更新机制。基于网络系统的实际问题,本文提出利用在线传送过来的态势时间序列数据对网络参数进行更新,以便于在实时观测网络态势数据的情况下建立更有效的 ILSTM 态势预测模型。

首先,合理地利用已有的历史数据,建立 LSTM 预测模型;其次,在收集实际在线数据时,利用建立的预测模型得到预测值;然后,使用下一个采样时间的新观测数据作为上一个采样时间的真实值,并将预测值与实际值之间的误差加到总体样本误差中;最后,利用误差最小化对模型参数进行迭代更新。利用提出的 LSTM 模型改进的方法对参数进行实时更新,随着在线数据的增多,模型得到的预测值将越来越精确。相比于传统的预测模型,所提模型更有利于网络系统的在线监测。一般情况下,ILSTM 的更新公式为:

$$\begin{cases} f_t=\sigma(W_f\times[h_{t-1},x_t]+b_f) \\ i_t=\sigma(W_i\times[h_{t-1},x_t]+b_i) \\ \tilde{C}_t=\tanh(W_C\times[h_{t-1},x_t]+b_C) \\ C_t=f_t\times C_{t-1}+i_t\times\tilde{C}_t \\ o_t=\sigma(W_o\times[h_{t-1},x_t]+b_o) \\ h_t=o_t\times\tanh(C_t) \end{cases} \tag{1}$$

4　NAWL 优化算法

4.1　改进的 Nadam 算法

Nadam(NGA Adaptive moment estimation)[14]是一种结合 NGA[15]和 Adam[16]算法的优化算法。Adam(Adaptive moment estimation)算法是一种具有自适应学习率的随机优化方法,融合了 AdaGrad[17]和 RMSProp[18]算法的优势。

ILSTM 模型中可训练的权重分别为 4 个层的权重矩阵,将不同层的权重矩阵连接起来,得到矩阵$[W_f,W_i,W_C,W_o]$,该矩阵将作为 NAWL 优化算法的输入,在训练中对优化算法的权重矩阵进行优化微调。4 个权重矩阵的大小分别为 $V\times H$,$H\times H$,$V\times H$,$C\times H$,因此可训练的权重的总数为 $(2\times V+H+C)\times H$。通常,V 的数量级高于 H 和 C,权重的总数可能达到 10^8,甚至更多。为了降低训练成本,并提高 Nadam 的收敛速度,本文利用 AWL 算法[19]中的 Look-ahead 方法对 Nadam 算法进行改进,得到了具有 Look-ahead 的 Nadam(Nadam with Look-ahead,NAWL)算法。为了有效地实现 NAWL,本文将使用一个等价公式,并引入一个新的超参数来控制 Look-ahead 强度。

算法 1　NAWL

Require:η 为学习速率

Require:μ_t,ν 在 0～1 之间,为矩估计的指数衰减率

Require:f(x)为代价函数

Require:$\theta=[W_f,W_i,W_C,W_o]$ 为优化的参数向量

$m_0\leftarrow 0$(初始化一阶矩参数向量)

$n_0\leftarrow 0$(初始化二阶矩参数向量)

$t\leftarrow 0$(初始化时间步长)

$z_0\leftarrow 0$(初始化中间变量)

While θ_t not converged do

$t\leftarrow t+1$

$g_t\leftarrow\nabla_{x_{t-1}}f(x_{t-1})$(获得前瞻参数的梯度)

$\hat{g}_t\leftarrow\dfrac{g_t}{1-\prod\limits_{i=1}^{t}u_i}$

$m_t\leftarrow\mu_t\cdot m_{t-1}+(1-\mu_t)\cdot g_t$(更新一阶矩估计)

$\hat{m}_t\leftarrow\dfrac{m_t}{1-\prod\limits_{i=1}^{t+1}\mu_i}$

$\bar{m}_t\leftarrow(1-\mu_t)\hat{g}_t+\mu_{t+1}\hat{m}_t$

$n_t\leftarrow\nu\cdot n_{t-1}+(1-\nu)\cdot g_t^2$(更新一阶矩估计)

$\hat{n}_t\leftarrow\dfrac{n_t}{1-\nu^t}$

$z_t\leftarrow\dfrac{\bar{m}_t}{\sqrt{\hat{n}_t}+\varepsilon}$(更新中间变量)

$\Delta\boldsymbol{\theta}\leftarrow-\eta[(1+\gamma)z_t-\gamma z_{t-1}]$(更新参数向量)

end while

Return θ_t

算法 1 中,ε 是为了避免除数为 0 的平滑常数项,通常取值为 1×10^{-8};$\nu=0.999$,$\mu_0=0.99$ 表示矩估计的指数衰减率;$\gamma=0.9$ 是引入的新的超参数,用来控制前瞻强度。

通过上述 NAWL 的更新公式,根据迭代次数对 ILSTM 模型的权重$[W_f,W_i,W_C,W_o]$进行更新。可以看出,NAWL 是使用 Look-ahead 方法对 Nadam 算法进行的改进,其继承了 AWL 的 Look-ahead 方法和 Nadam 超强的收敛速度的特性,不仅降低了 Nadam 的训练成本,还提高了 Adam 和 Nadam 的收敛速度,其对 ILSTM 网络模型的收敛速度的提升要远远优于 Adam 和 RMSProp 等经典的优化算法。

4.2　NAWL 算法的收敛分析

引理 1　由文献[10]可得:Nadam 的参数更新公式中的 $\dfrac{\bar{m}_t}{\sqrt{\hat{n}_t}+\varepsilon}$ 是渐近收敛的,即 Nadam 是渐近收敛的。

引理 1 的证明过程请参考文献[10]。

定理 1　Nadam 算法具有渐近收敛性,则 NAWL 算法也具有渐近收敛性。

证明:假设学习速率不随时间变化,则可得每次 NAWL 的参数更新为:

$$\Delta\boldsymbol{\theta}=-\eta[(1+\gamma)z_t-\gamma z_{t-1}] \tag{2}$$

对参数更新式(2),有:

$$\lim_{t\to\infty}(\sum_{t=0}^{n}-\eta((1+\gamma)\cdot z_t-\gamma\cdot z_{t-1}))$$

$$=\lim_{t\to\infty}\eta(\gamma\cdot z_n-\gamma\cdot z_0)+\sum_{t=0}^{n}\eta\cdot z_t \tag{3}$$

由引理可知,$z_t=\frac{\bar{m}_t}{\sqrt{\hat{n}_t}+\varepsilon}$是渐近收敛的。由式(3)可知,如果$\lim_{t\to\infty}\alpha(\gamma\cdot z_n-\gamma\cdot z_0)$渐近收敛,则 NAWL 算法渐近收敛。因为 $z_0=0$ 和 η 和 γ 都是恒定的常数,而且根据 Nadam 执行训练优化时,z_n 接近于 0,因此:

$$\lim_{t\to\infty}\mu\cdot(\gamma\cdot z_n-\gamma\cdot z_0)=\lim_{t\to\infty}\eta\cdot\gamma\cdot z_n\approx 0 \tag{4}$$

由此可得:式(3)是渐近收敛的,即 NAWL 算法具有渐近收敛性。

5 基于 NAWL-ILSTM 的态势预测模型

本文主要利用在线更新机制对 LSTM 网络参数进行改进,并且使用 Look-ahead 方法对 Nadam 进行改进,得到的 NAWL 优化算法提高了收敛速度。本文提出的基于 NAWL-ILSTM 态势预测算法的步骤如下。

步骤 1 实际时间序列为 $X=(x_1,x_2,\cdots,x_n)$,将时间序列 X 扩展成矩阵$\begin{bmatrix}x_1 & x_2 & \cdots & x_{n-k+1}\\ x_2 & x_3 & \cdots & x_{n-k+2}\\ \vdots & & & \\ x_k & x_{k+1} & \cdots & x_n\end{bmatrix}$,其中 n 为时间序列的长度,k 为样本数量。训练样本被表示为 $y=(x_k,x_{k+1},\cdots,x_n)$。利用式(5)对时间序列 X 进行标准化:

$$X=\frac{x_i}{\sqrt{x_i^2+x_{i+1}^2+\cdots+x_{i-k+1}^2}},i=1,2,\cdots,n-k+1 \tag{5}$$

步骤 2 初始化网络参数并设置超参数。

$$\begin{cases}W_f=rand(L,N)\\ b_f=rand(1,N)\\ \quad\vdots\\ Max\_iter=M_1\\ Error\_Cost=M_2\end{cases} \tag{6}$$

其中,M_1 和 M_2 分别代表误差阈值 $Error\_Cost$ 最大迭代次数 Max_iter 和 L 为 LSTM 细胞单元数;N 为神经元层数;W_f 和 b_f 分别为忘记门权重和偏置。类似地,有输入门和输出门等。

步骤 3 计算需要忘记的细胞单元状态信息:

$$\hat{f}_t=\sigma(W_f\cdot[h_{t-1},x_t]+b_f)*C_{t-1} \tag{7}$$

计算出遗忘门的输出后,将遗忘门的输出和前一时刻的单元状态相乘。

步骤 4 计算 t 时刻可以保持在细胞单元状态中的信息。

$$\hat{i}_t=\sigma(W_i\cdot[h_{t-1},x_t]+b_i)*\tanh(W_C\cdot[h_{t-1},x_t]+b_C) \tag{8}$$

式(8)包括两个部分,第一部分是输入门 i_t 的输出,它决定了细胞单元需要更新的值;第二部分是利用 tanh 函数来创建新的候选向量 C_t。然后,将候选向量与输入门的输出相乘。

步骤 5 计算出细胞单元状态 C_t:

$$C_t=\hat{i}_t+\hat{f}_t \tag{9}$$

细胞单元状态是输入门和遗忘门状态结合的结果。

步骤 6 计算 t 时刻的网络输出:

$$h_t=\sigma(W_o\cdot[h_{t-1},x_t]+b_o)\cdot\tanh(C_t) \tag{10}$$

首先,计算输出门 O_t,再将 O_t 和当前时刻的单元状态相乘,得到当前时刻的网络输出。h_t 为当前时刻的预测值。重复步骤 3-步骤 6,计算出所有训练样本的预测值。

步骤 7 计算所有预测值 h 和真实值 y 的误差:

$$J_{(\theta)}(y,h;W,b)=\frac{1}{2}\|y-h\|^2 \tag{11}$$

使用 BPTT 算法进行反向传播以更新网络参数,迭代次数为 1,然后转到步骤 3,若达到误差阈值或最大迭代次数,即 $error>Error\_Cost$ 或者 $iter>Max\_iter$,则退出训练循环。

步骤 8 输入待更新的权重矩阵 $\boldsymbol{\theta}_0=[W_f,W_i,W_C,W_o]$,并利用 NAWL 算法对 ILSTM 网络模型的参数进行训练。

$$\theta_t=\theta_{t-1}-\eta[(1+\gamma)z_t-\gamma z_{t-1}] \tag{12}$$

由于参数初始化是添加新样本时历史样本的全局最优解,因此只需执行几个循环步骤就可以实现新样本下的全局最优解。

步骤 9 根据在线观测数据实时更新参数方法,再添加新样本 $X_{n+1}(x_{n-k+2},\cdots,x_{n+1})$ 和 $\boldsymbol{\theta}_0$,进行步骤 3-步骤 6 的前向传播,并且得到新样本的预测值 h_{n+1}。

$$error=error+\frac{1}{2}(h_{n+1}-x_{n+2})^2 \tag{13}$$

步骤 10 当下一采样时刻的预测值达到网络被攻击点时,网络安全管理员会发出网络被攻击的警告并迅速做出反应,以防止网络被进一步攻击。

由此可得,NAWL-ILSTM 态势时间序列预测算法的流程如图 4 所示。

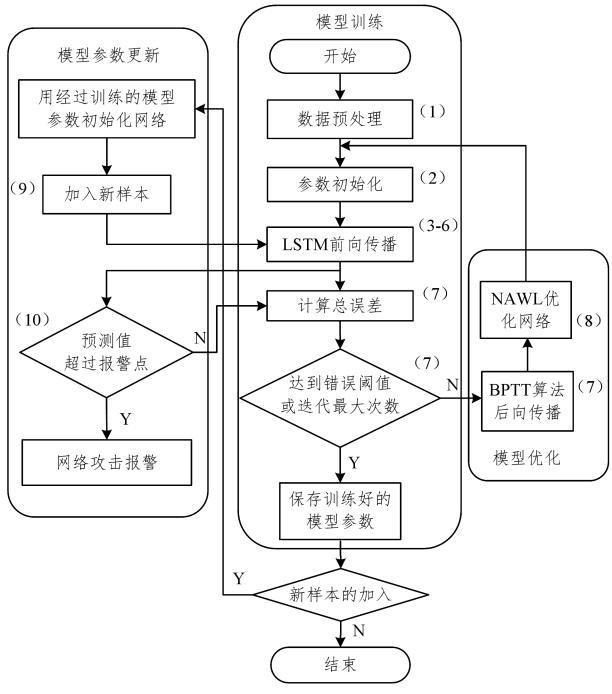

图 4 基于 NAWL-ILSTM 的态势序列预测算法的流程图

Fig. 4 Flow chart of situation time series prediction based on NAWL-ILSTM

6　实验与结果分析

6.1　态势时间序列预处理

为了验证 NAWL-ILSTM 预测方法的有效性,采用某网络公司 7 月到 9 月中 95 天的防火墙、IDS 等收集到的历史日志信息作为原始数据集进行实验,每天对日志信息样本采集一次。将前 77 天的数据作为训练集,后 18 天的数据作为预测集,并用上述评估方法把原始数进行量化计算,从而得到网络安全态势值。

由于安全态势值是随机的,且量纲差异大,为提升本文方法的训练速度,用上述方法对原始数据进行标准化处理,得到的网络安全态势时间序列如图 5 所示。

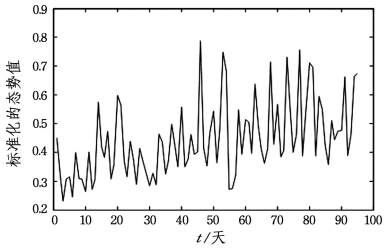

图 5　标准化的网络安全态势时间序列

Fig. 5　Standardized network security situation time series

6.2　实验结果的分析与比较

6.2.1　复杂度分析

按照本文算法步骤分析时间复杂度,可以得到 $O(K(H+CS+(H+3SC)I))=O(KW)$,其中 K 为输出单元的数量,C 为存储单元块的数量,S 为存储单元块的大小,H 为隐藏层数量,I 为与记忆细胞、门单元、隐藏单元连接的前向连接单元数量,W 为权重且 $W\leqslant K(H+CS)+(H+CS+2C)I$。上面的权重值是通过计算输出单位在权重方面的所有导数而获得的。其中,$H+SC$ 为直接连接输出单元的数量;CSI 为连接记忆细胞的数量;HI 为连接隐藏层的数量;$2CI$ 为连接门单元的数量。由于单个门单元影响记忆细胞 S,因此通过链式法则对块大小进行求和,计算出通向门单元的所有输出单元的导数,其复杂度为 $2CIS$。从而可以得到结论:通过给定 N 个记忆细胞,可以计算出 ILSTM 算法的复杂度为 $O(N^2)$。将本文算法与其他预测方法进行复杂度对比,结果如表 1 所列,其中 T 是 HMM 的状态数,D 是问题维度。

表 1　不同算法的复杂度比较

Table 1　Complexity comparison of different algorithms

算法	复杂度
HMM	$O(T^2\times D)$
SVM	$O(D\times D)$
RBF	$O(N\times D\times g_{\max})$
ILSTM	$O(N^2)$

6.2.2　收敛性分析

Nadam 吸收了 Look-ahead 方法和 Nadam 算法的特性,利用了 NGA 算法中梯度下降的优势来实现 Adam 优化算法对网络权重的更新,加快了模型的收敛速度;利用 Look-ahead 方法对 Nadam 进行改进,使得训练成本大大降低,更加快了收敛速度。从图 6 可以看到,不同的优化算法对网络训练的收敛速度有不同程度的提高,其中 NAWL 所需的迭代次数为 60,是最少的,没有进行改变的 Nadam 优化算法迭代次数达到了 80,Adam 和 RMSProp 算法甚至迭代了 100 次还未完全收敛。因此,NAWL 优化算法提高了 ILSTM 网络训练的收敛速度,优化了学习效率。

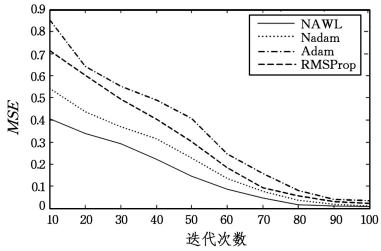

图 6　均方误差随迭代次数的变化曲线

Fig. 6　Change curve of MSE with iterative times

6.2.3　预测精度对比

为了体现所提方法的优势,选取未被改进的 LSTM,RBF,SVM 这 3 种最常用的预测模型进行对比,结果如图 7 所示。可以看出,所有的预测方法都能很好地处理历史数据,但是几种传统方法的预测效果并不好。本文方法由于使用更多的在线数据来更新 ILSTM 预测模型的参数,因此其预测值相比于其他预测方法更接近实际值,而其他方法与实际值均有不同程度的误差。本文方法可以很好地利用在线数据来提高预测精度,实验结果验证了该算法的有效性。

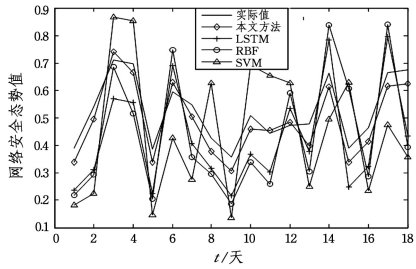

图 7　不同算法预测的态势值对比

Fig. 7　Comparison of situation prediction of different algorithms

为了能够更加整体地评价不同预测模型的预测性能,从均方误差和平均相对误差两个方面进行对比分析,结果如图 8 所示。从图中可以发现,相比于传统的 LSTM,RBF 和 SVM,本文改进的 LSTM 预测模型可以得到更趋近于真实网络的结果,且预测误差相对较小。这说明文中提出的预测模型具有更好的预测性能,且更加接近实际的网络安全态势时间序列。

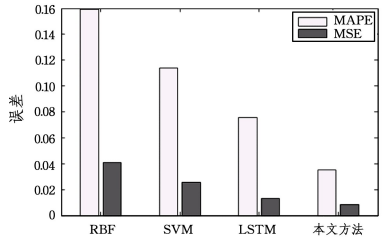

图 8　不同预测模型的误差对比

Fig. 8　Error comparison of different prediction models

结束语　本文提出了一种基于 NAWL 算法优化在线更

新机制 LSTM 的网络安全态势预测方法。仿真结果表明,较其他预测模型而言,该方法在历史数据较少的情况下可以达到更准确的预测效果,很好地拟合了网络安全的发展趋势。与其他 3 种预测模型相比,本文方法能够更好地处理长时间序列不确定性的问题,且具有优越的预测性能;通过与不同的优化算法进行比较,本文采用 Look-ahead 方法改进了 Nadam,提高了收敛速度,减少了 ILSTM 神经网络模型的学习时间。下一步主要对 ILSTM 神经网络模型进行改进和稳定性分析,以达到进一步提高预测精度的效果。

参 考 文 献

[1] JAJODIA S,LIU P,SWARUP V,et al. Cyber Situational Awareness: Issues and Research [M]. Boston, MA: Spring-Verlag,US,2010.

[2] BOX G E P,JENKINS G M,REINSEL G C. Time series analysis forecasting and control,4th Edition[M]. Beijing: Posts & Telecom Press,2005:19-180.

[3] LIANG W,CHEN Z,YAN X,et al. Multiscale Entropy-Based Weighted Hidden Markov Network Security Situation Prediction Model[C]// IEEE International Congress on Internet of Things. IEEE,2017:97-104.

[4] LI F W,ZHENG B,ZHU J,et al. A method of network security situation prediction based on AC-RBF neural network[J]. Journal of Chongqing University of Posts & Telecommunications,2014,26(5):576-581. (in Chinese)
李方伟,郑波,朱江,等. 一种基于 AC-RBF 神经网络的网络安全态势预测方法[J]. 重庆邮电大学学报(自然科学版),2014,26(05):576-581.

[5] JIANG Y,LI C H,YU L S,et al. On Network Security Situation Prediction Based on RBF Neural Network[C]// 2017 36th Chinese Control Conference,Beijing: Technical Committee on Control Theory of Chinese Association of Automation, 2017: 4060-4063.

[6] ZHANG S M,LI B X,WANG B Y. The Application of an Improved Integration Algorithm of Support Vector Machine to the Prediction of Network Security Situation[J]. Applied Mechanics & Materials,2014,513-517(513-517):2285-2288.

[7] DUAN M. Short-Time Prediction of Traffic Flow Based on PSO Optimized SVM[C]// International Conference on Intelligent Transportation,Big Data & Smart City. IEEE Computer Society,2018:41-45.

[8] WANG X,WU J,LIU C,et al. Fault time series prediction based on LSTM cyclic neural network [J]. Journal of Beijing University of Aeronautics and Astronautics,2018,44(4):772-784. (in Chinese)
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.

[9] CHEN Z,LIU Y,LIU S. Mechanical State Prediction Based on LSTM Neural Netwok[C]// China Control Conference,Beijing: Technical Committee on Control Theory of Chinese Association of Automation. 2017:3876-3881.

[10] ZHU J,MING Y,SONG Y H,et al. Mechanism of situation element acquisition based on deep auto-encoder network in wireless sensor networks[J]. International Journal of Distributed Sensor Networks,2017,13(3):155014771769962.

[11] LING F W,ZHANG X Y,ZHU J,et al. Network security situation assessment model based on information fusion [J]. Journal of Computer Applications,2015,35(7):1882-1887. (in Chinese)
李方伟,张新跃,朱江,等. 基于信息融合的网络安全态势评估模型[J]. 计算机应用,2015,35(7):1882-1887.

[12] SUN R Q. Research on the price trend prediction model of the stock index based on LSTM neural network[D]. Beijing:Capital University of Economics and Business,2016. (in Chinese)
孙瑞奇. 基于 LSTM 神经网络的美股股指价格趋势预测模型的研究[D]. 北京:首都经济贸易大学,2016.

[13] GREFF K,SRIVASTAVA R K,KOUTNIK J,et al. LSTM: A Search Space Odyssey[J]. IEEE Transactions on Neural Networks & Learning Systems,2015,28(10):2222-2232.

[14] DOZAT T. Incorporating Nesterov Momentum into Adam[R]. Stanford University,2015.

[15] SUTSKEVER I,MARTENS J,DAHL G,et al. On the importance of initialization and momentum in deep learning[C]// International Conference on International Conference on Machine Learning. JMLR. org,2013:1139-1147.

[16] BALLES L,HENNING P. Dissecting Adam:the sign,magnitude and variance of stochastic gradients[C]// International Conference on Machine Learning. New York: ACM,2018:693-709.

[17] DUCHI J,HAZAN E,SINGER Y. Adaptive Subgradient Methods for Online Learning and Stochastic Optimization[J]. Journal of Machine Learning Research,2011,12(7):257-269.

[18] YEUNG S,RUSSAKOVSKY O,NING J,et al. Every Moment Counts:Dense Detailed Labeling of Actions in Complex Videos [J]. International Journal of Computer Vision,2017(8):1-15.

[19] ZHANG C,ZHANG C,ZHANG C. An improved Adam Algorithm using look-ahead[C]// International Conference on Deep Learning Technologies. New York: ACM,2017:19-22.