

混合多层结构化网络的攻击路径预测算法研究

王婷婷¹, 李晓东¹, 任 刚²

(1 郑州成功财经学院 信息工程系, 河南 郑州 451200 ; 2 中国科学院 软件研究所, 北京 100190)

摘 要: 针对目前的攻击路径预测方法不能准确反映攻击力对后续路径的影响, 提出了混合多层结构化网络的攻击路径预测算法。首先, 对攻击路径样本时间序列进行组合排序, 重新组合产生攻击路径样本, 采用极限学习机进行分类训练, 然后采用人工鱼群捕鱼算法进行学习机优化, 提高对网络攻击路径样本的全局搜索和局部寻优能力, 避免陷入局部最优解。最后得到的仿真结果表明, 采用该方法进行运动员攻击路径预测的收敛性较好, 迭代次数较低, 预测精度得到提升。

关键词: 混合多层; 捕鱼算法; 极限学习机; 攻击路径; 预测

中图分类号: TP311

文献标识码: A

文章编号: 1000-7180(2017)04-0131-04

Research on Attack Path Prediction Algorithm for Hybrid Multilayer Structured Networks

WANG Ting-ting¹, LI Xiao-dong¹, REN Gang²

(1 Department of Information and Engineering, Zhengzhou Chengong University of Finance and Economics, Zhengzhou 451200, China;

2 Institute of Software, Chinese Academy of Sciences, Beijing 100190, China)

Abstract: For the current attack path prediction method can not accurately reflect the impact of attack power on the subsequent path, an attack path prediction algorithm is proposed. First of all, the combination of the sort of attack path sample time series, re combination of attack path samples, using extreme learning machine classifier, and then use artificial fish fishing machine learning algorithm was optimized to improve the global search for network attack path samples and local optimization ability and avoid the local optimal solution. The simulation results show that the proposed method has better convergence, lower iteration number and higher prediction accuracy.

Key words: hybrid multilayer; fishing algorithm; extreme learning machine; attack path; prediction

1 引言

攻击路径预测是建立在对原始采集的攻击路径数据的特征分析和信息处理基础上的, 通过构建攻击路径的非线性时间序列分析模型, 对原始的攻击路径数据信息流进行样本特征空间组合排序与重组, 提取攻击路径序列的非线性特征, 实现攻击路径预测, 取得了较好的成果, 典型的算法比如极限学习机算法^[1]、模拟退火算法和支持向量机预测算法等^[2-3], 上述算法在对攻击路径的先验样本知识缺乏和可用数据信息特征较少的情况下预测精度不高,

对此, 本文提出一种捕鱼算法优化极限学习机的攻击路径预测模型。

2 攻击路径样本时间序列进行组合排序

为了实现对混合多层结构化网络的攻击路径的准确预测, 首先对攻击路径样本时间序列进行组合排序, 通过前期的统计分析和攻击路径采样, 统计得到攻击路径数据信息流的单变量时间序列表示 $\{x_n\}$, 其中攻击路径数据采样样本长度为 N , 攻击路径数据时间序列的信息流模型描述为

$$x_n = x(t_0 + n\Delta t) = h[z(t_0 + n\Delta t)] + \omega_n \quad (1)$$

式中, $h(\cdot)$ 为攻击路径预测模型的传递函数, ω_n 为攻击路径数据采样的统计误差, 采用无监督学习算法进行攻击路径样本时间序列的统计误差修正, 定义 k 时刻统计样本 i 向样本 j 的转移概率:

$$P_{ij}(k) = \frac{\hat{w}_{ij}^p(k)/d_{ij}^p(k)}{\sum \hat{w}_{ij}^p(k)/d_{ij}^p(k)} \quad (2)$$

式中, $\hat{w}_{ij}(k)$ 为样本时间序列分布空间之间的权值差, $d_{ij}(k) = 1/\eta_{ij}(k)$ 为样本序列的相轨迹之间的欧式距离, 在攻击路径数据预测过程中, 权值之差越大, 相轨迹之间的间距越大, 转移概率越大, 由此得到攻击路径样本时间序列组合排序的重构轨迹为

$$X = [s_1, s_2, \dots, s_K]_n = (x_n, x_{n-\tau}, \dots, x_{n-(m-1)\tau}) \quad (3)$$

式中, $K = N - (m-1)\tau$ 表示攻击路径数据时间序列的攻击路径数据相似度(相关性), m 为信息重组的嵌入维数, 用 $s_i = (x_i, x_{i+\tau}, \dots, x_{i+(m-1)\tau})^T$ 表示单组攻击路径样本的标量信息, 由此实现运攻击路径样本时间序列组合排序, 重新排列的序列为

$$r_n = y(x_n) = y_{\text{rank}(x_n)}, \quad n = 1, 2, \dots, N \quad (4)$$

对重排后的攻击路径样本 $\{r(n)\}_{n=1}^N$ 作 Fourier 变换得到新序列 $y'(n)$, 按照新序列的秩进行自回归移动, 从而有效反应原始攻击路径时间序列的静态属性.

3 基于极限学习机的攻击路径测量误差修正

采用极限学习机对攻击路径样本时间序列组合排序结果进行分类训练, 实现攻击路径测量误差修正, 攻击路径测量误差的优化目标问题的数学描述如下:

$$\begin{aligned} \min F(x) &= [f_1(x), f_2(x), \dots, f_n(x)] \\ \text{s. t. } g_i(x) &\leq 0 (\text{或 } \geq 0) \quad i = 1, 2, \dots, n \\ h_j(x) &= 0, \quad j = 1, 2, \dots, m \end{aligned} \quad (5)$$

式中, $f_i(x) (i = 1, 2, \dots, n)$ 为攻击路径测量误差修正的目标函数; $g_i(x)$ 为不等式约束; $h_j(x)$ 为相关性统计约束条件. 这里, 引入基于极限学习机的攻击路径测量误差修正多目标优化中几个重要概念^[4]:

定义 1 极限学习机支配集: 攻击路径预测的决策变量 x^* 支配 x 满足: 所有 $f_i(x^*) \leq f_i(x)$, 并且至少存在一个 $f_i(x^*) < f_i(x)$, 其中 $i, j = 1, 2, \dots, n$, 此时的极限学习机支配集满足局部收敛.

定义 2 Pareto 最优解: 对于攻击路径预测的判别统计量 $X^* \in S$, 当且仅当存在一个弱非线性可

行解 $X \in S$, 使得所有不等式 $f_i(X^*) \leq f_i(X)$ 成立, 其中 $i = 1, 2, \dots, n$, 在感知邻域范围内存在一个 i , 使得严格不等式 $f_i(X^*) < f_i(X)$ 成立, 此时的攻击路径预测的统计量 X^* 为多目标优化问题, 通过求得 Pareto 最优解, 能满足预测模型的收敛性.

根据上述定义, 给出基于极限学习机的攻击路径测量误差修正的步骤描述如下所示.

Step1: 初始化极限学习机的参量, 极限学习机中冗余向量集的位置 Pbest. 每个种群的个体最佳位置 Pbest 由下面的公式确定:

$$\begin{aligned} V_{ij}(g+1) &= V_{ij}(g) + c_1 r_{1ij}(g)[Pbest_{ij}(g) - x_{ij}(g)] + \\ &\quad c_2 r_{2ij}(g)[Gbest_j(g) - x_{ij}(g)]; \end{aligned}$$

Step2: 设定极限学习机的输入节点个数 N 和向量元素个数 k , 攻击路径的输入向量:

$$x(t) = (x_0(t), x_1(t), \dots, x_{k-1}(t))^T \quad (6)$$

加权向量:

$$Gbest_j(g+1) = \arg \min_{Pbest_{ij}} f(Pbest_{ij}(g+1)) \quad (7)$$

为了平衡极限学习机全局搜索能力和局部搜索能力, 采用惯性权重 w 进行误差修正自适应寻优, 寻优公式如下:

$$\begin{aligned} V_{ij}(g+1) &= wV_{ij}(g) + c_1 r_{1ij}(g)[Pbest_{ij}(g) - x_{ij}(g)] + c_2 r_{2ij}(g)[Gbest_j(g) - x_{ij}(g)], \\ 0 &\leq w \leq 1 \end{aligned} \quad (8)$$

Step3: 更新极限学习机的非劣解, 实现最小迭代次数下有限收敛, 调整权重向量 α ;

Step4: 更新极限学习机中种群的局部最优位置, 攻击路径测量误差修正的判决统计量表达:

$$GD = \frac{\sqrt{\sum_{i=1}^n d_i^2}}{n} \quad (9)$$

当 $GD = 0$ 时, 得到攻击路径测量的误差修正公式如下:

$$DM = \frac{d_e + d_b + \sum_{i=1}^{n-1} \left| d_i - \frac{\sum_{i=1}^{n-1} d_i}{n-1} \right|}{d_e + d_b + (n-1) \frac{\sum_{i=1}^{n-1} d_i}{n-1}} \quad (10)$$

式中, d_e 为集合 S_s 中的极值点, d_b 为边界点到非劣解的距离. 通过上述步骤进行极限学习机的攻击路径测量误差, 避免预测结果陷入局部最优解.

4 攻击路径预测算法模型优化

在上述进行攻击路径的时间序列分析和测量

误差修正的基础上,进行攻击路径预测模型优化设计,本文提出采用人工鱼群算法进行极限学习机优化,避免预测值陷入局部最优解,人工鱼群算法是模拟鱼群体的觅食行为^[5-6],通过个体写作的最优选择进行自适应寻优的仿生算法,假设 D 维搜索空间中人工鱼当前状态为 X_i ,其感知邻域范围内的鱼群个体的目为 N_f ,进行极限学习机优化的目标函数为

$$C(n) = \min \left\{ \sum_{i=1}^n [STA(i) - ETA(i)] \right\} \quad (11)$$

在捕鱼算法设计中,人工鱼状态 X_i 和 X_j 之间的距离表示攻击路径的两个特征量之间的欧式距离,描述为

$$l(X_i, X_j) = \|X_i - X_j\| \quad (12)$$

上述的距离表示在攻击路径预测中的无监督学习的差异化程度,通过降低鱼群的拥挤度,提高局部收敛能力,需要优化权值子集 $\{W_O\}_{i=1}^{N-m-a}$,鱼群的子集 $\{W_{\text{final}}\} = \{\{W_H\}, \{W_C\}, \{W_O\}\}$. 优化后的觅食伙伴中心位置的鱼群分布子集可以表示为

$$\{W_O\}_{i=1}^{N-m-a} = \{\{x_O^i\}_{i=1}^{N-m-a}\} \quad (13)$$

若 $(N_f / N) < \delta$,用 visual 表示人工鱼的感知范围,visual 表示捕鱼的试探次数,初始化人工鱼群的当前状态^[7],在试探 trynumber 次后,得到优化组合后的攻击路径样本属性集记为

$$\begin{aligned} & \text{minimize } \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \\ & \text{subject to } y_i - (w \Phi(x_i) + b) \leq \varepsilon - \xi_i \quad (14) \\ & (w \Phi(x_i) + b) - y_i \leq \varepsilon - \xi_i^* \\ & \xi_i, \xi_i^* \geq 0, i = 1, 2, \dots, n; C > 0 \end{aligned}$$

式中, ξ_i 和 ξ_i^* 非负松弛变量, C 和 ε 为人工鱼群的核极限学习最好位置,是全局极值,得到捕鱼算法优化后的攻击路径预测的最终差别函数为

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(x_i, x_j) + b \quad (15)$$

式中, α_i and α_i^* 是由惩罚因子绑定,极限学习机的混合核函数满足 Mercer 条件^[8],在捕鱼过程中,鱼群个体间的协作来寻找最优解,得到攻击路径预测的自适应加权表达式为

$$K_{\min} = \beta K_{\text{poly}} + (1 - \beta) K_{\text{rbf}}, \beta \in (0, 1) \quad (16)$$

式中, $K_{\text{poly}} = [(x \cdot x_i) + 1]^2$ 表示错分样本惩罚程度控制核函数, $K_{\text{rbf}} = \exp(-\gamma \|x - x_i\|^2)$ 表示初始鱼群中每条人工鱼的初始状态适应度核函数, β 是调整鱼群适应度的权系数. 若达到鱼群最大迭代次数,得到攻击路径预测的最优值将其赋给目标鱼

群,实现预测算法改进,伪代码表述:

Swarm Behavior;

if $(\frac{N_f}{N} < \delta \& \& Y_c < Y_i)$

$$X_{i \rightarrow \text{next}} = X_i + \text{rand}() \cdot \text{step} \cdot \frac{X_c - X_i}{\|X_c - X_i\|}$$

else goto search

(17)

根据上述描述,得到基于捕鱼算法和极限学习机的攻击路径预测模型的实现流程如图 1 所示.

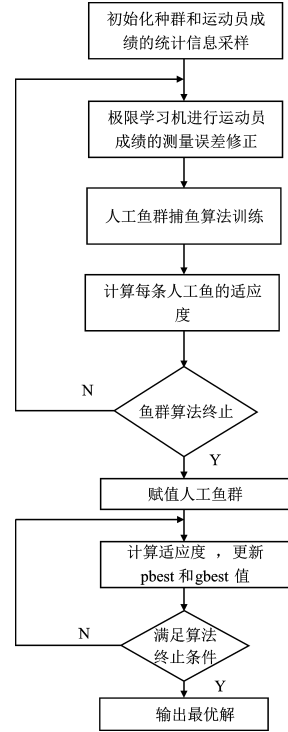


图 1 攻击路径预测模型设计流程图

5 实验测试分析

对攻击路径预测的仿真实验环境建立在 Matlab7.10 仿真软件上,实验的硬件环境为主频 3.0 GHz,内存 2 GB 的计算机,仿真参数取 $W_{\min} = 0.4$, $W_{\max} = 0.9$, $C_{\min} = 1.5$, $C_{\max} = 2.0$,对攻击路径的样本数据进行归一化处理: $z_i = \frac{x_i}{\max(x_i)}$, $i = 1, 2, \dots, 5$,人工鱼群训练参数对 $(\beta, C, \gamma, \varepsilon)$ 的搜索范围分别为 $[0, 1]$ 和 $[0.1, 100]$,极限学习机的最优参数对 $(\beta, C, \gamma, \varepsilon) = (0.0817, 2.7854, 1.0268, 0.0069)$,交叉概率 $0.1 \rightarrow 0.3$,测试函数的参数设置见表 1.

根据上述仿真参量设定,进行攻击路径预测,预测训练中鱼群种群规模为 100,得到采用本文方法

和传统方法的收敛曲线如图 2 所示。

表 1 测试函数的参数设置

参数	种群规模	分布性度量值
KDW1	200	14 000
KDW2	300	14 000
KDW3	300	15 000
KDW4	100	15 000
KDW5	100	14 000
KDW6	100	12 000

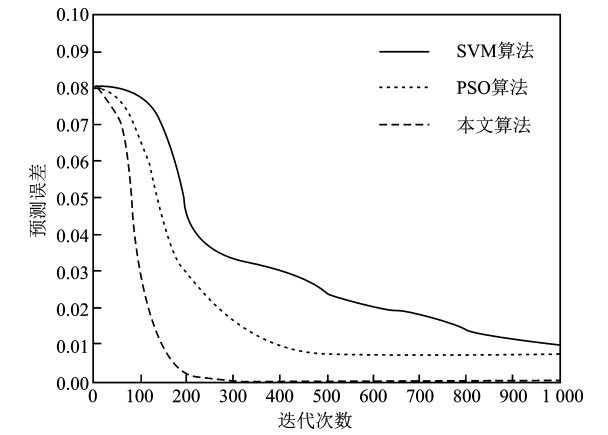


图 2 攻击路径预测收敛曲线

分析图 2 结果得知,采用本文算法进行攻击路径预测,能在最短的迭代次数下实现预测误差收敛,在迭代 230 次后误差收敛到 0,而采用传统的 SVM 算法和 PSO 算法,误差无法收敛到零,且迭代次数较多,容易陷入局部收敛.表 2 给出了采用不同算法进行攻击路径预测的精确度、运行时间以及均方根误差对比结果,分析得知,本文方法的预测精度较高,误差最小,消耗时间也较少。

表 2 性能对比

预测方法	PSO	SVM	PSO-RBFNN	本文捕 鱼算法
错误预测个数	3	4	6	0—
预测准确率/%	58.333	70.43	66.634	100
均方误差(MSE)	4.234	0.653 2	0.234 482	0.004 655
运行时间/s	4.976	1.528 7	4.306	0.590

6 结束语

本文提出一种采用人工鱼群算法进行极限学习机优化的攻击路径预测模型改进设计,提高攻击路径预测准确度,避免预测值陷入局部最优解.对攻击

路径采样统计样本时间序列进行组合排序,重新组合产生攻击路径样本,采用极限学习机进行分类训练,采用人工鱼群捕鱼算法进行学习机优化,提高对攻击路径样本的全局搜索和局部寻优能力,避免陷入局部最优解.研究证明,采用该方法进行攻击路径预测的误差较低,收敛性较好,耗时较少。

参考文献:

[1] 陈迎春,孙昱,莫有权,等.基于结构和参数智能优化的神经网络数值预测方法研究[J].信息工程大学学报,2016,17(5):23-24.

[2] XYLOMENOS G,VERVERIDIS C N,SIRIS V A,et al.A survey of information-centric networking research[J].IEEE Communications Surveys Tutorials,2014,16(2):1024-1049.

[3] PIRO G,GRIECO L A,BOGGIA G,et al.Information-centric networking and multimedia services: present and future challenges[J].Transactions on Emerging Telecommunications Technologies,2014,25(4):392-406.

[4] 吕慧,赵跃华.基于掩码的 AES 算法抗二阶 DPA 攻击方法研究[J].电子设计工程,2015,23(20):31-33.

[5] 金俊尉,楼佩煌,钱晓明,等.机床横梁结构神经网络与算法优化研究[J].机械科学与技术,2016,35(9):1359-1364.

[6] 江务学.基于结构优化递归神经网络的网络流量预测[J].西南大学学报(自然科学版),2016,38(2):149-154.

[7] 徐倩漪,齐芳.基于机器学习的通信网络非结构化大数据分析算法研究[J].激光杂志,2016,33(10):125-128.

[8] 王玲玲,孙艳丽,王康.基于激光传感器的路径识别算法与实现[J].电子设计工程,2016,24(4):26-28.

作者简介:

王婷婷 女,(1981-),硕士研究生,讲师.研究方向为计算机网络、无线传感器网络.

李晓东(通讯作者) 男,(1980-),硕士,副教授.研究方向为软件工程、计算机网络.

E-mail:lixiaodong@chenggong.edu.cn.

任刚 男,(1978-),博士研究生,讲师.研究方向为并行软件与计算科学.