文章编号:1671-4598(2023)09-0267-07 DOI:10.16526/j.cnki.11—4762/tp.2023.09.039 中图分类号:TP393.08 文献标识码:A

基于数据聚类的网络安全防护态势优化方法

李 星[1],李浩然[2]
(1.武警特色医学中心 信息科,天津 300162;
2.中国科学院天津工业生物技术研究所 生物设计中心,天津 300308)

摘要:针对传统模糊特征检测方法存在的效率低、精度不高等问题,设计了一种新的网络安全防护态势优化模型;对网络安全状态分布进行建模,并利用数据挖掘技术对网络信息进行挖掘;利用新型入侵识别检测方法对所设计的网络安全估计状态进行自适应特征提取,提取网络安全状况的特征数据集和处理单元;采用模糊C平均数据聚类方法(FCM)提取综合信息;对入侵特征信息流进行分类,根据属性分类结果进行网络安全态势预测,实现安全态势评估;基于不同场景下进行实验,结果表明,所提算法适用于网络安全的场景,准确性和鲁棒性都得到了验证。

关键词:数据聚类;网络安全防护;预测;数据挖掘

Optimization Method of Network Security Protection Situation Based on Data Clustering

LI Xing[1], LI Haoran[2]
(1. Characteristic Medical Center of Chinese People's Armed Police Force, Information Centre, Tianjin 300162, China;
2. Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China)

Abstract: Aiming at the problems of low efficiency and low accuracy of traditional fuzzy feature detection methods, a new network security protection situation optimization model is designed. The distribution of network security state is modeled, and the data mining technology is used to mine network information. The new intrusion identification detection method is used to extract adaptive features from the designed network security estimation state, and extract the characteristic data set and processing unit of network security state. Fuzzy C mean data clustering method (FCM) is used to extract the comprehensive information. The intrusion characteristic information flow is classified, and the network security situation is predicted according to the attribute classification results, and the security situation evaluation is realized. Based on the experiments in different scenarios, the results show that the proposed algorithm is suitable for network security scenarios, and its accuracy and robustness are verified.

Keywords: data clustering; network security protection; prediction; data mining

0 引言

在目前人们日常应用网络的环境中,总是存在着许多网络安全风险[1-7]。这主要是因为随着互联网、大数据、云计算、人工智能等技术的快速发展和应用,使得网络空间安全面临的风险和威胁日益增多。DDoS攻击、APT攻击、高危漏洞增多、数据暴露事件频发、“灰色”应用层出不穷、高新技术带来的安全风险等问题尤为突出。目前,网络系统面临的安全问题主要有以下几个方面:涉及的网络安全数据量逐渐增加,规模越来越大;网络安全事件不断地碎片化,使其难以被察觉,而获得的安全信息分散无序,管理员需要花费大量的时间和精力来分析潜在的安全威胁,既费时又费力,事半功倍;而现有的许多网络安全系统在数据采集上存在局限性,有的仅局限于网络安全数据的一个或几个方面的采集、分析和处理,难以全面描述和反映网络安全状况。面对这些新的挑战和威胁,现有的传统网络安全防御手段、策略和方法(如入侵检测系统、防火墙、反病毒、访问控制等)已不能跟上当今网络系统的实际安全需求。由于当前网络入侵逐渐呈现规模化、隐蔽性的特点,导致常规的网络安全评估、监测和防护模型已不能满足需求,因而利用数据整合与有效的数据分析手段构建更加可靠的网络安全态势评估模型已成为研究的重点之一。

态势感知技术由Endsley于1988年提出,定义为“在一定时空条件下对环境因素的识别和理解,并预测未来趋

收稿日期:2023-04-08; **修回日期**:2023-04-25。
作者简介:李 星(1989-),女,硕士,助理工程师。
引用格式:李 星,李浩然.基于数据聚类的网络安全防护态势优化方法[J].计算机测量与控制,2023,31(9):267-273.

势”。他将情境感知模型分为三层：情境抽取、情境理解和情境预测。文献［8］提出了一种新型网络安全态势模型，设计了一种结合多源数据的网络安全态势架构。在此模型的基础上，定义了网络安全态势。网络安全态势评估就是指在各类网络数据中筛选出有效信息，然后输入至对应的评估模型，并计算得到对应的态势数值，然后根据数值评价出此时网络的安全状态，从而为之后可以提前预测及防护提供支撑及参考。因此网络安全评估方法是当前网络的安全状况防护关键技术之一，有利于提升安全防护效率。

态势感知的思想最早出现在军事领域，用于确定军事环境和态势，后来应用于交通、医疗等领域，并延伸到网络安全领域[9-15]。网络安全态势感知是指通过一系列技术收集尽可能多的网络安全要素，建立相应的评估和预测模型，帮助网络管理者及时应对风险。网络安全态势预测是指根据历史态势评估数据预测网络未来的状态，防止网络攻击。预测的前提是相邻数据点之间存在一定的规则。研究表明，网络流量数据具有自相似性。网络安全态势预测的对象是网络安全态势值，网络态势值是按时间顺序排列的，相邻数据点之间存在一定的规律。因此，网络安全形势具有可预见性和可行性。

目前，网络安全形势预测主要采用的方法有以下几种。(1) 自回归移动平均模型：它是基于一个平滑的时间序列预测未来状态，但它对时间序列的长度有一定的要求。(2) 灰色理论：侧重于灰色关联来发现系统的内在规律，但对波动较大的数据预测效果较差；(3) 时间序列：它基于相邻数据之间的相关性，但在建模过程中需要考虑很多元素。(4) 神经网络：使用安全事件作为输入和输出的态势值，实现网络安全态势预测，但容易陷入局部收敛，影响预测效果。

对于大多数深度学习从业者来说，序列建模就是循环网络的同义词。然而，最近的研究结果表明，卷积架构在音频合成和工业生产力等任务上优于循环网络。变压器已应用于自然语言处理、计算机视觉、语音识别等领域。构建了结合 Transformer 和 TCN 的网络安全态势预测模型，并以 UNSW-NB15 和 CSE-CIC-IDS2018 为基准数据集，通过对比实验验证了模型的有效性。

在最新的复杂网络安全评价中使用多源数据融合技术，不但能够提高评价的准确度，而且能够提高互联网应用的安全性。为有效的运用数据整合方法以适应多源数据结构分析性能的需要，很多研究者提出了安全态势评价模型[5-10]。其模型由服务器、网络和用户三层组成。在服务器层，根据受到入侵和威胁信息的主机系统评估模型，在服务器层的入侵信息得到融合后，模型将数据发送到网络层。以入侵数据为参考对象，对存在危险特性的网络信息进行了分类，并对网络信息系统中的可能存在的危险特性和风险相关信息进行建模和评价，得到整体结果。

1 网络安全态势的研究现状

网络安全态势感知根据网络所处的当前环境因素确定网络态势，从而预测网络近期的状态。网络安全态势预测是网络态势感知的重要组成部分。它可以尽快识别网络中潜在的安全风险，并充分评估这些潜在威胁的影响程度，帮助网络安全管理者掌握当前网络状况，以便在网络攻击发生之前对这些威胁采取遏制和预防措施[16-20]。

近年来，许多学者结合各种技术，提出了各自的网络安全态势感知模型。机器学习在模型构建中的应用，极大地提高了数据挖掘的准确性和效率。态势预测作为态势感知的重要组成部分，在实际模型构建过程中，经常通过网络安全时间序列对未来网络安全态势进行实验，但需要更大的数据集和存储容量作为支撑。

文献［21］提出了一种基于异构传感器事件流的多阶段网络攻击态势实时感知方法，首次建立了基于网络连通性和攻击过程语义的攻击建模方法，生成多级攻击模板。然后将实时警报流中的攻击事件语义与攻击模板进行关联，完成多阶段攻击的态势感知。

文献［22］针对无线网络环境提出了一套更容易实现的网络流量安全指标。通过收集和可视化无线网络中的数据包到达间隔时间等指标，帮助网络管理人员识别攻击，以掌握网络情况。文献［23］提出了一种警报关联框架，可以有效地检测多步攻击事件并预测攻击者行为。

循环神经网络（RNN）作为一种时间序列分析模型，在处理非线性关系方面有较好的表现。因此，它被广泛应用于不同领域的时间序列预测任务。与长短期记忆(LSTM) 相比，传统 RNN 存在梯度消失问题，在长期依赖预测问题中表现较差。为了克服 RNN 的局限性和梯度消失问题，提出了 RNN、LSTM 和门控循环单元（GRU）的渐进模型[24]，基于这两种模型的编解码器在机器翻译中都取得了良好的效果。然而，随着序列长度的增加，它们的性能会迅速下降，为了解决这一问题，提出了基于注意机制的编码器－解码器网络。

现有的大部分网络安全攻击预测方法基本都是个体预测工具，这可能会带来几个相应的问题：(1) 个体预测工具神经元数量较少，对参数设置更敏感，存在过度训练的问题；(2) 单个探测器的预测精度不稳定，没有标准精度可用作比较；(3) 与集成学习（融合机器学习算法）相比，单个检测器的预测精度有限；(4) 目前大多数架构都是“黑盒”模型。模型内的参数以非线性的方式进行交互控制，我们无法捕捉到，模型参数的调整过于复杂和不可确定。

为了解决上述问题，许多研究人员开始通过选择新模型或组合新模型来寻找新的解决方案。目前的研究是在序列建模任务上对卷积和循环架构进行的最广泛的系统比较。结果表明，序列建模和循环网络之间的共同联系应该重新考虑。TCN 结构不仅比 LSTM 和 GRU 等典型循环网络更

精确，而且更简单、更清晰。TCN 不使用门控机制，具有较长的内存[25]。因此，这可能是将深度网络应用于序列处理的一个更合适的起点。文献［26］开发了一个使用 TCN 对时间序列建模的轻量级预测系统，并且提出了一种基于信道注意力的时间卷积模型，以较少的参数实现了卫星图像时间序列分类。

上述模型在短期预测任务中表现较好，但网络安全态势预测需要有多步时间序列预测的能力，以满足长序列处理需求。Transformer 模型在捕获长期依赖关系方面表现优于 RNN 模型。它对自注意机制的依赖可以有效地避免圆形结构。基于自我关注的 Transformer 模型最近在翻译、音乐和图像生成方面表现更好。然而，自注意性的空间复杂度随着序列长度的增加呈二次增长，这对超长序列的处理造成了计算能力的限制。网络安全数据是具有细粒度和长期相关性的时间序列。因此，单一模型已经不能满足长期预测需要解决的问题。文献［27］构建了 TCN-LSTM 混合模型，实现了网络安全态势预测。文献［28］利用时序融合模型实现了可解释的高性能多视图预测。通过时间融合变压器的选择和相关分量的抑制，得到不同尺度下的时间关系。

文献［29］针对传统极限学习机在网络安全态势预测方法中存在的预测精度低、收敛速度慢等问题，提出一种基于改进麻雀算法优化极限学习机的预测方法，提高了网络安全态势预测的准确性。文献［30］针对当前网络安全态势评估模型存在的准确性低，收敛慢等问题，提出了一种融合模拟退火、麻雀算法和 BP 神经网络的网络安全态势评价模型。但目前许多方法对于综合复杂网络背景下防护态势评估的准确性都不佳，需要进行优化和调整。

现有的时间序列预测方法依赖于机器学习模型的自动特征选择来识别相关变量，同时支持基于多个时间序列和测量的安全情况预测。现有的研究文章大多应用多个模型架构的组合，每个组合的模型也会有特定的适用领域和相应的精度。总体而言，网络安全态势感知相关成果较多，但其相关技术尚处于发展阶段。目前提出的各种算法模型各有优缺点。大多数模型没有全面考虑安全态势影响因素，态势评估和预测的计算精度有进一步提高的空间。

针对上述问题，提出了一种新的网络安全防护态势优化方法。建立了网络安全状态模型，并引入数据挖掘技术对各类网络安全信息进行挖掘。利用入侵检测方法提取自适应特征和主要的功能参数，进而提取敏感信息。然后利用优化后的 FCM 方法对复杂高频信息流进行分类。通过实验验证了所提方法的有效性。

2　模型分析与预处理

2.1　复杂场景下网络入侵的安全状态模型

为了进一步实现最优的评价性能，利用信号处理框架设计了综合评价算法。数据是一组广泛而稳定的非线性时间信号模型，适用于复杂的网络环境。

IPoE 业务是一种接入认证业务。在 IPoE 业务中，用户通过物理以太网链路访问，通过 DHC 动态获取 IP 地址。

用户认证主要有三种类型：web 认证、option 60 认证和线路认证。通过引入“累积状态变化”，值导数 GRU 算法可以同时定量描述移动网络恶意流量的低层和高层变化信息。此外，通过增加池化层，值导数 GRU 算法可以获得关于流的重要信息。病毒入侵给网络安全带来隐患。通过对参数的详细讨论，利用式（1）表示病毒入侵信息流的特征。网络环境下 m 终端上病毒入侵信息流的特征具体表达式如下：

$$x(k)=[x_1(k),x_2(k),\cdots x_m(k)]\quad i=1,2,\cdots,m \tag{1}$$

式中，k 为网络安全态势分布的属性值，x_i（k）为网络入侵的特征标量的时间序列。假设 n 维随机分布变量（x_1，x_2，…，x_n）受网络病毒攻击，表达式如下：

$$\begin{cases} v_s=\left\|X_s-\sum\limits_{i=1}^{n}\omega_i X_i\right\|^2, & s=1,\cdots,n \\ \omega_s=\dfrac{\dfrac{1}{\dfrac{v_s}{\sum\limits_{j=1}^{n}v_j}+\lambda}}{\sum\limits_{j=1}^{n}\dfrac{1}{\dfrac{v_s}{\sum\limits_{j}^{n}v_j}+\lambda}}, & s=1,\cdots,n \end{cases} \tag{2}$$

式中，v_s 为网络收集网络安全入侵中数据的变化行为。利用网络入侵行为所收集的数据中 x_s 和 r_t 的偏差，使得复杂场景下网络入侵模型的状态为 $V=\{V_1,V_2,\cdots,V_n\}$，则此时网络安全状况的条件转移概率 L 表示为：

$$L=C-m\sum_{s=1}^{n}\log(\sigma_s)-\sum_{s=1}^{n}\sum_{t=1}^{m}\frac{(x_{st}-r_t)^2}{2\sigma_s^2} \tag{3}$$

式中，C 为网络安全入侵的免疫值，σ_s 表示从所收集数据中任意选取的状态向量，$a_{i,j}$ 表示网络安全态势分布状态 i 在空间 j 上的分布概率。网络入侵检测的稳态概率表达式如下：

$$\sigma_s^2=\frac{1}{m}\sum_{t=1}^{m}(x_{st}-r_t)^2 \tag{4}$$

在信息融合中心设置网络病毒攻击的平均互通信息权重属性如下：

$$\Phi(\omega_1,\omega_2,\cdots\omega_n)=E\{\exp[j(\omega_1x_1+\cdots+\omega_nx_n)]\} \tag{5}$$

网络安全威胁特征的振幅和频率估计如下：

$$m_k=E[x^k]=\int_{-\infty}^{+\infty}x^kf(x)\mathrm{d}x \tag{6}$$

$$u_kE[(x-n)^k]=\int_{-\infty}^{+\infty}(x-n)^kf(x)\mathrm{d}x \tag{7}$$

改进的计算方法表达式如下：

$$\theta_1(k+1)=\theta_1(k)-\mu\mathrm{Re}[y(k)\varphi^*(k)] \tag{8}$$

节点分布位置图如图 1 所示。构建网络安全入侵的安全状态分布模型。当有黑客试图入侵网络节点时，其不仅会考虑当前节点的价值，还会考虑入侵代价和可能产生的收益。入侵的代价和收益不会改变节点之间的初始状态变

换，但会对入侵节点的选取产生影响，黑客会倾向于选取代价低、收益大的节点进行入侵。

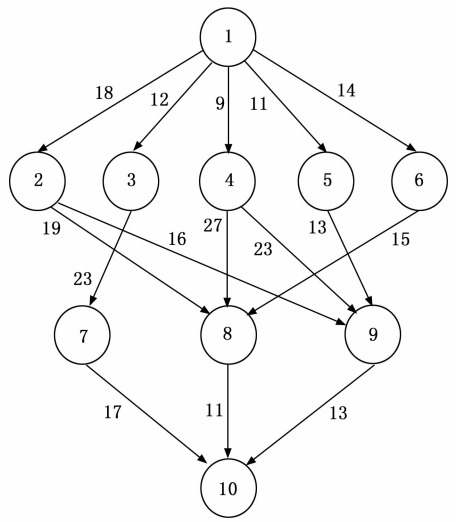

图 1 网络安全态势节点分布图

2.2 网络安全态势关联信息模型的构建

ARMA 模型用于模拟网络攻击环境中影响网络安全状况的威胁指数和主机威胁指数。本文引入强化学习（RL）进行模型搭建，RL 大多基于实际场景进行学习，并不是提前继续数据，因此本文利用真实动态场景作为仿真环境。先在样本集随机选择一个新的样本，并针对分类器的预期目标进行奖励，随后再根据学习目标对算法进行初始化，从而通过环境提高分类器的预测难度。可以计算黑客对当前节点发起攻击的概率。当攻击概率为 0 时，表示攻击无法获益，此时黑客不会对节点进行攻击；当概率为 1 时，表示攻击行为可以获益，此时黑客将会发起攻击。

因此，不但要考虑网络用户的静态数据，而且还要考虑移动网络用户的动态变化数据。通过收集有关移动网络用户的静态和动态数据，从而提升算法对网络空间中危险信息的检测精度。

网络安全态势威胁指数可表示为：

$$\begin{cases} x_k = f(x_{k-1}) + v_k \\ y_k = h(x_k) + e_k \end{cases} \tag{9}$$

式中，v_k，e_k 表示为时空差偏差特征流，则对安全态势指数的威胁表示为：

$$p(\omega_k) = t_{\tilde{v}_k}(\tilde{u}_k, \tilde{\Sigma}_k) = \frac{\Gamma\left(\frac{\tilde{v}_k + d}{2}\right)}{\Gamma\left(\frac{\tilde{v}_k}{2}\right)} \frac{|\tilde{\Sigma}_k|^{-\frac{1}{2}}}{(\tilde{v}_k \pi)^{d/2}} \tag{10}$$

式中，$\tilde{u}_k$ 为动态全局概率搜索状态向量，$\tilde{\Sigma}_k$ 为待解变量，$\Gamma(\cdot)$ 为 Sigma 函数。在数据聚类的约束下，得到网络入侵交叉变异的概率为：

$$x(n) = s(n) + v(n) = \sum_{i=1}^{L} A_i \cos(\omega_i n + \varphi_i) + \sum_{j=0}^{\infty} h(j) w(n-j) \tag{11}$$

入侵特征分为（w_1，w_2，…，w_n），n 为预测误差。在此基础上，利用数据聚类提取网络攻击特征，实现安全态势评估。

在真实的网络场景中，网络安全指标会根据网络的运行状态而动态变化。当黑客的目标已知时，静态风险评价的精确度也会随之下降。因此，可以利用基于 Bayes 原理得出的动态可达率，通过时刻更新网络节点的可达率，建立动态风险评价模型。考虑以下两种修改：

1）根据 SDN 问题模型模拟强化学习的环境，该环境的状态为网络入侵类型。

2）Agent 是复杂的分类器，其主要任务是通过模拟环境的状态预测流量的类别。

3 数据聚类和网络安全防护态势评估

3.1 基于数据聚类的安全态势特征检测

为了构建动态场景环境下的安全状态分布安全威胁，假设输入网络安全估计模型的自适应全局概率分布为 $x(t)$，并使用属性分类结果。网络安全态势的范围和频率估计如下：

$$W_x(t,v) = \int_{-\infty}^{+\infty} x(t+\tau/2) x^*(t-\tau/2) e^{-j2\pi v\tau} \mathrm{d}\tau \tag{12}$$

$$W_x(t,v) = \int_{-\infty}^{+\infty} X(v+\xi/2) X^*(v-\xi/2) e^{-j2\pi\xi t} \mathrm{d}\xi \tag{13}$$

式中，W_x（t，v）为匹配范围内数值交换的入侵数据的脉冲响应，为一实数。该问题的最优解决方案是在数据聚类中找到最优个体。基于自适应数据分类定义模型：

$$E_x = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} W_x(t,v)\mathrm{d}t\mathrm{d}v \tag{14}$$

FCM 数据提取入侵特征信息流，将服务器层的网络入侵数据分解为数据聚类特征。得到的交叉概率为：

$$\begin{cases} \int_{-\infty}^{+\infty} W_x(t,v)\mathrm{d}t = |X(v)|^2 \\ \int_{-\infty}^{+\infty} W_x(T,v)\mathrm{d}v = |x(t)|^2 \end{cases} \tag{15}$$

病毒数据的跨项目分布特征描述如下：

$$x(k) = [x_1(k), x_2(k), \cdots, x_m(k)] \quad i = 1,2,\cdots,m \tag{16}$$

受到网络安全威胁的网络的用户特征定义为：

$$\begin{cases} y(t) = x(t-t_0) \Rightarrow w_y(t,v) = W_x(t-t_0, v) \\ y(t) = x(t) e^{f2\pi v_0} \Rightarrow W_y(t,v) = W_z(t, v-v_0) \end{cases} \tag{17}$$

对于所有 ω，$|V(e^{j\omega})| = 1$，选择一个集合适应度函数，使网络安全态势检测方法的频响模量在 $z = e^{\pm j\omega_0}$，保证了算法的收敛性。

3.2 网络安全防护态势评估

结合数据聚类算法检测网络病毒攻击的信息流，通过在整个搜索空间中的时频展开，将模型的经验模态分布指向性函数定义为：

$$y(t) = \sqrt{k} x(kt), k > 0 \tag{18}$$

$$W_y(t,v) = W_x(kt, v/k) \tag{19}$$

根据所建立的成本计算模型，得出各数据聚类中心对

应的成本。通过数据聚类的特征约束，网络安全态势分析的时频响应应为：

$$y(t)=\int_{-\infty}^{+\infty}h(t-s)x(s)\mathrm{d}s \tag{20}$$

$$W_y(t,v)=\int_{-\infty}^{+\infty}W_k(t-s,v)W_s(s,v)\mathrm{d}s \tag{21}$$

如果得到的适应度较大，则通过数据聚类来测量病毒的攻击强度，得到网络安全态势评估的迭代方程为：

$$K_{k+1/k+1}=\rho_k\frac{m_k}{m_{k+1}}\Phi_k K_{k/k}\Phi_k^T+\frac{1}{m_{k+1}}\delta K \tag{22}$$

$$\delta K=\begin{bmatrix}S-\sigma I & z\\ z^T & \sigma\end{bmatrix} \tag{23}$$

$$m_k=\sum_{i=1}^{n}a_i,m_{k+1}=m_k+\delta m_{k+1},\sigma=\frac{1}{m_k}\sum_{i=1}^{n}a_i b_i^T r_i,$$

$$B=\frac{1}{m_k}\sum_{i=1}^{n}a_i b_i r_i^T,S=B+B^T \tag{24}$$

经调频得到的评价安全态势信号的模糊约束匹配输出如下：

$$R_s^{(0)}=\sum_{n=0}^{k}R_s^{(n)},d_{\gamma n}d_{\gamma n}+R_s^{(k+1)} \tag{25}$$

在此基础上，通过使用数据聚类约束的的一般分析方法，对模型的调幅信号进行检验，将网络入侵信息的两个交叉点所涉及的范围设置为匹配范围，并引入特征自相关变量 S，采用数据聚类对提取的网络入侵信息流进行自关联检验，从而完成对网络安全的精确评价。而不同的安全基本单元指标的特征参数，往往有着不同的维数和物理含义。如果将上述基本信息数据放入网络态势预测进行计算，物理单元中的各类数据会发生难以预见的偏差，从而使得无法成功预测网络态势。某样本适应度值越大，其可以成为样本集中心的概率就会越大。其中适应度值表示样本与当前聚类中心欧氏距离的最小值。

4　仿真实验分析

4.1　数据集介绍

选择 UNSW-NB15 数据集作为本文的数据集。UNSW-NB15 数据集是由澳大利亚网络安全中心（ACCS）网络边缘实验室的 IXIA PerfectStorm 工具创建的。UNSW-NB15 数据集是基于一个全面的网络环境设计的，用于生成攻击活动。该数据集从真实的、正常运行的网络中收集攻击数据集，满足网络安全态势预测需要使用具有时间特征和连续时间维的数据集的条件。数据集还提供了训练集和测试集，减少了数据预处理的工作量。UNSW-NB15 作为基准数据集，包含 Tcpdump 工作者捕获的 100 GB 原始流量。数据集包含 9 种类型的网络攻击，实施的攻击类型包括 FTP、SSH、DoS、Heartbleach、Web 攻击、渗透、僵尸网络和 DDoS 等。

网络流量数据通常用高维向量表示。采用 t 分布随机邻域嵌入（t-SNE）方法对其复杂度进行可视化，并基于可视化图对其进行定性分析。在 UNSW-NB15 数据集上呈现显著差异，其中一些类内距离可能大于类间距离，并且分布不均匀。正常样本和攻击样本具有相同的空间特征，这也说明特征空间不可能线性分离。因此，基于这个数据集实现网络安全态势预测，可以最大限度地模拟真实网络的复杂性。

由于 UNSW-NB15 数据集的连通性特征和其他特征，该数据集在攻击模式识别和分析方面具有巨大潜力。虽然使用 UNSW-NB15 数据集检测了预测模型的性能，但也发现了此数据集在在研究中的一些局限性。在进行实验之前，对数据集中的大量数据进行预处理，发现数据集中包含大量的噪声，这些噪声对情况预测的贡献很小。

4.2　数据预处理

在深入研究 Snort 威胁分类机制的基础上，我们首先将威胁级别分为高、中、低三类。第一类是侵入计算机并获得计算机控制权的攻击，可以对计算机系统造成致命威胁，定义为高。第二类攻击是为了获取系统内部的私人信息而进入计算机的攻击，这种攻击被定义为中。第三类攻击不进入计算机系统，目的是消耗网络带宽。这种类型的攻击使计算机无法与外界通信或提供正常的操作，它被定义为中。第四类是网络扫描型攻击，对计算机的影响较小，定义为低。

将权重系数理论与攻击威胁等级分类有机地结合起来，确定攻击威胁值。基于威胁等级越高威胁值越高的原理，对威胁等级进行了预测使用权重系数分布函数在 0 和 1 之间。具体表达式如下：

$$M_i=\begin{cases}\dfrac{1}{2}+\dfrac{\sqrt{-2\ln\dfrac{2i}{n}}}{6}, & 1\leqslant i<\dfrac{n}{2}\\ \dfrac{1}{2}, & i=\dfrac{n}{2}\\ \dfrac{1}{2}-\dfrac{\sqrt{-2\ln\left(2-\dfrac{2i}{n}\right)}}{6}, & \dfrac{n}{2}<i<n\end{cases} \tag{26}$$

式中，最大量化值定义为 $M_0=1$；n 表示威胁级数；i 表示威胁级别的序数，即 $i=0$，1 和 2 分别表示高、中和低。UNSW-NB15 数据集中各种攻击类型的威胁等级及威胁值如表 1 所示。

表 1　UNSW-NB15 数据集的攻击态势值

类型	威胁等级	威胁值
Analysis	中	0.55
Backdoor	高	1
Dos	中	0.65
Exploit	高	0.85
Fuzzers	高	1
Generic	中	0.65
Reconnaissance	低	0.35
Shellcode	高	1
Worm	高	0.90

为评估提出的模型的预测能力的准确性，使用均方根误差（*RMSE*）和平均绝对误差（*MAE*）来衡量情况预测准确性。*RMSE* 具体表达式如下：

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \tag{27}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N} | \hat{y}_i - y_i | \tag{28}$$

式中，$\hat{y}_i$ 代表预测值，y_i 代表实际值。这两个度量可以描述预测情况值和实际情况值之间的差异。在本文中，选择这两个指标来评估实验结果。

4.3 实验分析

本节将对所提出的模型进行验证。硬件设置为笔记本电脑，配置参数为 AMD R9-5800 CPU，运行内存 16 G，操作系统为 Windows 11。采用 MATLAB R2020b 编程软件进行算法设计。

网络入侵数据的采样尺度为 2 000，训练集为 120，模拟时间长度为 1 200 s，检测频率为 24 kHz。根据上述仿真回路和参数，进行网络安全防护态势评估仿真。首先，选择 UNSW-NB15 数据集上其中一段的含有强烈干扰信号的网络入侵行为数据为例，示意图如图 2 所示。以上述网络安全入侵信号为样本输入，作为评估模型的初始信息，从图 2 中可以看出，入侵数据受到媒体信息的干扰，难以有效识别一般。在防护态势评估中，以 8 s 的时间宽度提取模型的特征信息，然后通过数据聚类得到 Sink 节点和 Source 节点的网络，检测提取的网络安全威胁信息流与 Sink 节点和 Source 节点的网络之间的相关性。结果如图 3 所示。

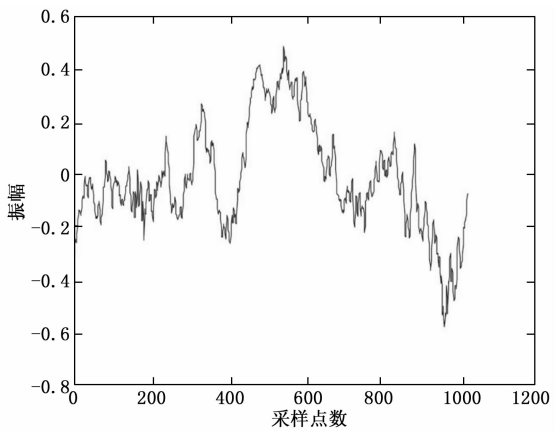

图 2 网络入侵信息数据

从图 2 和图 3 中可以看出，基于本文算法的网络安全态势评估具有良好的波束指向性，能够准确地反映出网络受网络安全威胁后对安全态势的最终分布情况，从而定量地比较本文方法在网络安全态势评估中的优越性能，本文模型与其他方法相比，以评价的准确性为检验指标，对比结果如图 4 所示。在迭代后期，本文模型较快的实现了防护准确率 100%的目标，另外两种算法前期准确率较低均低于

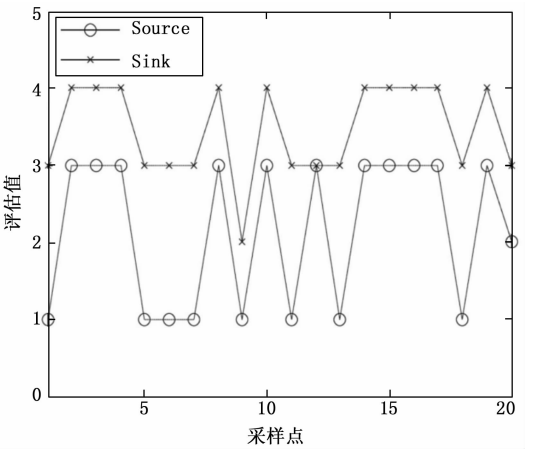

图 3 网络态势评估结果

75%且随着迭代的进行准确率提升较慢，因此本文算法具有更好的防护准确性及效率。

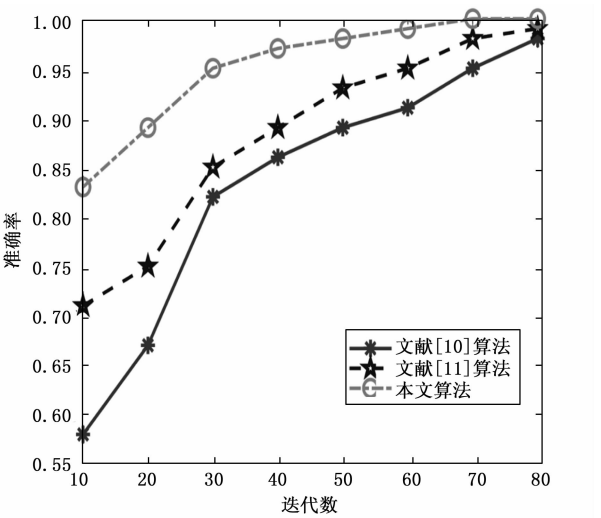

图 4 网络防护态势准确率比较

其次，在整个 UNSW-NB15 数据集上大规模进行实验，具体实验结果如表 2 所示。

表 2 UNSW-NB15 结果

方法	RMSE	MAE
文献[10]	0.186	0.152
文献[11]	0.152	0.111
文献[29]	0.131	0.094
本文算法	0.071	0.054

表 1 为各个算法在 UNSW-NB15 数据集上的 *RMSE*、*MAE* 结果，从中可以看出本文算法相比于文献［10］算法、文献［11］算法、文献［12］算法在 *RMSE* 减少了约 45.8%～61.8%，在 *MAE* 减少了约 42.5%～64.5%。展现出本文算法的优越性，表明本文算法误差更小，对网络安全态势的评估准确度更高，从而有利于实现网络安全态势

的优化。

5　结束语

本文设计了一种网络安全态势优化新方法。首先，构建多变场景下的网络安全状态趋势模型，利用综合大数据挖掘方法对网络安全信息相关数据进行挖掘，获得网络的综合状态和结构。其次，利用入侵检测方法提取自适应特征和主要功能参数结构，进而提取敏感信息。然后利用优化后的 FCM 方法对复杂高频信息流进行分类。仿真结果表明，本文算法能较好地进行网络安全防护状况评估，网络入侵检测的准确率较高，误差更小，网络安全得到了保证。在未来的研究中，可以尝试考虑模型在不同场景下的评估，以验证其性能。

参考文献：

[1] 童伟传. 基于 SimHash 算法的大数据网络安全态势的评估 [J]. 机械设计与制造工程，2022，51 (5)：125 - 129.

[2] 李峻屹. 基于大数据的 K-means 聚类算法的网络安全检测应用研究 [J]. 机械设计与制造工程，2021，50 (9)：115 - 118.

[3] 陆雨晶，陈　琳. 基于 FAHP 的网络安全态势感知风险评估技术研究 [J]. 计算机与数字工程，2021，49 (5)：957 - 960，976.

[4] 常利伟，刘秀娟，钱宇华，等. 基于卷积神经网络多源融合的网络安全态势感知模型 [J/OL]. 计算机科学：1 - 16 [2023 - 04 - 06]. http://kns.cnki.net/kcms/detail/50.1075.TP.20230214.1457.040.html

[5] 吴嘉竣，徐文辉. 基于 RBF 神经网络的电力信息网络安全态势感知 [J]. 自动化技术与应用，2022，41 (9)：103 - 105.

[6] 肖　鹏，王柯强，黄振林. 基于 IABC 和聚类优化 RBF 神经网络的电力信息网络安全态势评估 [J]. 智慧电力，2022，50 (6)：100 - 106.

[7] 李　文. 基于反向传播算法的网络安全态势信息识别——评《网络安全态势感知》 [J]. 中国科技论文，2021，16 (4)：467.

[8] 吕鹏鹏，王少影，周文芳，等. 基于进化神经网络的电力信息网安全态势量化方法 [J]. 计算机科学，2022，49 (S1)：588 - 593.

[9] 张　然，潘芷涵，尹毅峰，等. 基于 SAA-SSA-BPNN 的网络安全态势评估模型 [J]. 计算机工程与应用，2022，58 (11)：117 - 124.

[10] SUN H Z，WANG J，CHEN C，et al. ISSA-ELM：A Network Security Situation Prediction Model [J]. Electronics，2022，12 (1)：48 - 55.

[11] 钱　锦，徐汉麟. 面向建设多元融合高弹性电网的网络安全态势感知技术研究 [C] //浙江省电力学会，浙江省电力学会 2021 年度优秀论文集，中国电力出版社，2022：109 - 116.

[12] 廖天颖，杨斯博，窦润亮. 基于贝叶斯网络的大数据安全动态风险评估模型研究 [J]. 网络空间安全，2023，14 (1)：60 - 68.

[13] 肖军弼，华　力. 基于改进 SIR 模型的网络安全态势分析 [J]. 计算机系统应用，2023，32 (3)：48 - 57.

[14] 周金全，朱世伟，张建平. 基于大数据和人工智能的网络安全态势分析方法研究 [J]. 中国新通信，2022，24 (11)：111 - 113.

[15] 姚　卓. 基于实战化的集团企业网络安全主动防御技术研究与实践 [J]. 信息技术与网络安全，2022，41 (5)：25 - 31.

[16] 施驰乐. 电子政务系统网络安全安全防护之变——浅谈态势感知与安全运营平台 [J]. 中国信息化，2019 (6)：59 - 62.

[17] 孔　珍，孔　硕. 网络安全态势感知关键技术研究 [J]. 中国信息化，2022 (4)：60 - 62.

[18] 宋　欣，刘　娜，苏　叶. 基于态势感知的智慧图书馆网络安全防护体系构建 [J]. 中华医学图书情报杂志，2022，31 (3)：70 - 75.

[19] 周　云，刘月华. 基于深度强化学习的智能网络安全防护研究 [J]. 通信技术，2021，54 (11)：2545 - 2550.

[20] 王斌斌，徐培杰，王思蕊. 网络安全态势感知在集团型钢铁企业安全防护体系中的应用 [J]. 冶金自动化，2021，45 (S1)：244 - 250.

[21] 李　罡，王德超，邱　烨，等. 基于联合处置与智能分析的网络安全防护平台应用实践 [J]. 城市轨道交通，2021 (8)：51 - 54.

[22] 曹鲁喆. 基于深度学习的校园网络安全态势要素提取与评估方法研究 [D]. 北京：中国人民公安大学，2021.

[23] 孙建立，朱卫东，李歆丽. 高校网络安全服务新模式探索——以基于大数据分析与云端的网络安全综合服务平台为例 [J]. 中国教育信息化，2021 (3)：27 - 31.

[24] 粟　毅，安小伟，李振，等. 安全态势感知联动防护在地震网络安全保障的应用 [J]. 华南地震，2020，40 (4)：63 - 70.

[25] 冯　耀. 数据挖掘技术在网络安全态势分析中的应用 [J]. 电子技术与软件工程，2020 (7)：245 - 247.

[26] 魏凡翔. 面向网络安全感知的态势可视化技术研究 [D]. 南京：南京理工大学，2019.

[27] 张　晴. 基于隐马尔可夫模型的网络安全态势预测的研究 [J]. 网络安全技术与应用，2019 (3)：30 - 31.

[28] 符　睿. 基于 AI 的工业互联网网络安全态势平台设计 [J]. 网络安全技术与应用，2023 (1)：11 - 13.

[29] CHENG M，LI S M，WANG Y H，et al. A New Model for Network Security Situation Assessment of the Industrial Internet [J]. Computers，Materials & Continua，2023，75 (2)：58 - 69.

[30] 缪祥华，方绍敏. SDN 中基于 ACO-BP 神经网络的 DDoS 攻击检测方法 [J]. 数据通信，2022 (4)：42 - 46.