doi:10.3969/j.issn.1001-893x.2017.01.011
引用格式:李方伟,王森,朱江,等. 基于增强型概率神经网络的安全态势要素获取[J]. 电讯技术,2017,57(1):64-71. [LI Fangwei, WANG Sen, ZHU Jiang, et al. Security situation element acquisition based on enhanced probabilistic neural network[J]. Telecommunication Engineering, 2017,57(1):64-71.]

基于增强型概率神经网络的安全态势要素获取*

李方伟,王 森**,朱 江,张海波
(重庆邮电大学 移动通信技术重庆市重点实验室,重庆 400065)

摘 要:态势要素获取作为整个网络安全态势感知的基础,其质量的好坏将直接影响态势感知系统的性能。针对态势要素不易获取问题,提出了一种基于增强型概率神经网络的层次化框架态势要素获取方法。在该层次化获取框架中,利用主成分分析(PCA)对训练样本属性进行约简并对特殊属性编码融合处理,将其结果用于优化概率神经网络(PNN)结构,降低系统复杂度。以 PNN 作为基分类器,基分类器通过反复迭代、权重更替,然后加权融合处理形成最终的强多分类器。实验结果表明,该方案是有效的态势要素获取方法并且精确度达到 95.53%,明显优于同类算法,有较好的泛化能力。

关键词:网络安全;态势要素;数据处理;协同增强;概率神经网络
中图分类号:TN918.91;TP393 **文献标志码**:A **文章编号**:1001-893X(2017)01-0064-08

Security Situation Element Acquisition Based on Enhanced Probabilistic Neural Network

LI Fangwei, WANG Sen, ZHU Jiang, ZHANG Haibo
(Chongqing Key Laboratory of Mobile Communications Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

Abstract: Situation elements extraction is the basis of the whole network security situation awareness and its quality will directly affect the performance of the situation awareness system. To solve the problem that the situation element is difficult to extract, a method is proposed to extract the hierarchical frame situation elements based on the enhanced probabilistic neural network (PNN). In the hierarchical access frame, the principal component analysis (PCA) is used to reduct the training sample attribute and process the special attribute encoding fusion. The result is used to optimize the structure of PNN and reduce the system complexity. PNN is taken as the base classifier to form the final strong classifier by repeated iteration, weight replacement and weighted fusion. The experimental results show that the scheme is an effective method to obtain the situation factors and its accuracy is 95.53%, significantly better than other similar algorithms.

Key words: network security; situational factors; data handling; synergistic enhancement; probabilistic neural network (PNN)

1 引 言

随着网络规模的快速发展,网络融合方式的多样化,无线传感器网络(Wireless Sensor Network, WSN)、P2P(Peer to Peer)、ad hoc 等多种网络组合,使网络拓扑结构更加复杂,网络资源频繁交互,通信量与流量急剧增加等。现有的安全技术如杀毒软件、防火墙、

* **收稿日期**:2016-05-09;**修回日期**:2016-09-07 Received date: 2016-05-09; Revised date: 2016-09-07
基金项目:国家自然科学基金资助项目(61271260);重庆市科委自然科学基金资助项目(cstc2015jcyjA40050);重庆市教委科学技术研究项目(KJ120530)
** **通信作者**:senma1990@163.com **Corresponding author**: senma1990@163.com

入侵检测系统都是单一的纵深防御机制,缺乏对网络整体的认识,难以向网络管理人员提供准确、实时的安全信息。在这种形势下,网络安全态势感知技术应运而生,并且得到了人们的广泛关注[1-2]。

网络安全态势感知是在动态的、复杂的网络安全环境中,从安全设备采集多源异构数据,通过数据融合、特征挖掘等技术进行整合,然后获取影响网络正常运行的安全态势要素,进一步理解,通过信息的融合处理和分析推理来描述网络当前安全状况,并对将来的发展趋势作出预测。作为一种自主的防御技术,它能有效地把不确定信息和异常信息提供给网络管理人员,帮助网络管理人员准确迅速地作出决策。态势要素获取作为整个网络安全态势感知的基础,是系统能够作出准确评估和预测的先决条件。

态势要素获取关键是在复杂的网络环境中判断网络中的每条数据信息是否存在异常,如果存在异常判断它属于网络异常中的哪一种异常,其核心就是分类问题,按照网络连接正常和异常的标准进行分类。不同于其他数据分类,态势要素获取是对多种网络设备产生的网络安全数据进行处理,不是根据某一属性进行分类,而是通过每条连接中多属性综合来判断类别,数据具有冗余、海量、高维度、异构和实时性等特点,并且最终的态势要素提供给系统作为态势评估和态势预测的基础。目前网络安全态势要素获取还处于初步研究阶段,相关成果较少,需要借助于已有的技术和理论。文献[3]基于海量的网络安全数据中难以发现入侵检测,建立了一种有选择的显示安全数据结合系统管理员认知获取态势要素的工具。文献[4] 针对军事战场中态势要素提取,提出一种基于概念的信息提取技术,用于解决态势感知中的信息提取问题。文献[5]将数据挖掘应用于多传感器态势感知框架中,通过融合、关联、统计等技术对态势要素进行提取。文献[6-7]提出将进化策略用于优化神经网络参数,提高了神经网络的收敛速度,然后用神经网络来对态势要素进行提取,提高分类精度,但是存在容易陷入局部最优的问题。文献[8]提出了逻辑回归和改进的粒子群优化算法(Logistic Regression and Improved Particle Swarm Optimization,LR-IPSO)模型,为了提升局部和全局搜索能力,采用非线性递减随机策略优化改进粒子群算法(Improved Particle Swarm Optimization,IPSO)权重值,提升逻辑回归(Logistic Regression,LR)模型的学习能力,是一种有效的态势要素获取方法,但是主要是针对小样本事件。网络态势要素信息收集是以传感器网络为主,由于传感器网络本身受限于能量和容量,因此加大了信息采集的难度。

为了使态势要素获取体系更加完备,本文建立了基于 Agent 的结构化网络安全态势要素获取框架,根据任务和功能的不同,该模型分为 3 层。利用主成分分析(Principal Component Analysis,PCA)对原始训练数据进行属性约简并对特殊属性进行编码融合处理,然后利用其结果对概率神经网络(Probabilistic neural network,PNN)结构网络进行优化,降低系统结构复杂度,减小数据传输过程中的消耗。为了快速、准确地获取态势要素,本文以优化后的 PNN 作为基分类器,然后经过反复迭代、权重更替,最后通过加权投票形成最终的强分类器。实验结果表明,本文设计的态势要素获取方法是有效的,为态势感知系统性能分析打下了良好的基础。

2　网络安全态势要素获取框架

Agent[9]是一种在分布式计算领域能持续主动引导发挥作用的计算实体,每个实体具有创建、调度、执行的过程,具有灵活性、竞争合作、自主适应等特点。多 Agent 系统在面对大型、复杂的问题时,各 Agent 依据一定的规则进行通信协作,能在网络内外进行统一部署,实现分布、自治的网络安全数据采集、分析和管理[10]。为了最大限度地降低每个节点获取态势信息的资源消耗,本文设计的模型根据节点任务和功能的不同,建立采集、分析、管理节点的层次化态势要素获取框架。其中,管理 Agent 作为根节点,在态势要素获取过程中,采集 Agent 周期性网络数据,并进行属性约简上传到上一层分析模块,直至上传到协调管理层,汇集整个网络的安全要素。框架如图 1 所示。

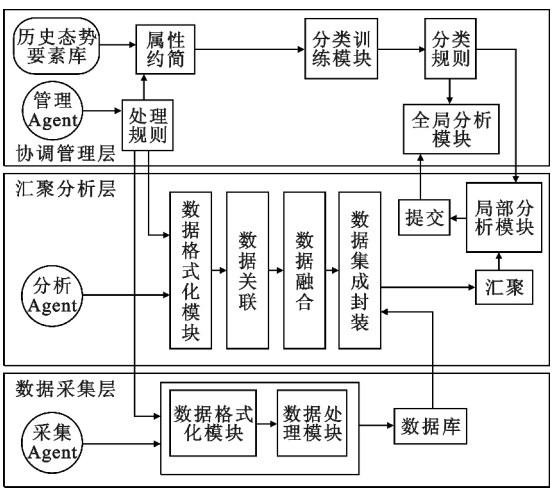

图 1　层次化安全态势要素获取框架

Fig1. A hierarchical security situation element acquisition framework

在数据采集层,通过各数据采集 Agent 收集安全信息,包括日志数据、数据包分析记录、漏洞扫描、终端监控等数据。经过格式转换、属性约简形成数据库上传到汇聚分析层。

汇聚分析层由各分析 Agent 组成,同样采集整个网络的流量信息、拓扑信息等数据,和数据采集层上传的信息一起进行格式统一化、关联、编码融合、封装后,进入局部分析模块处理后产生局部态势要素,最后提交局部态势要素给协调管理层。

协调管理层由各管理 Agent 组成,各管理 Agent 向数据采集层和汇聚分析层分发处理规则、分类规则并进行动态管理调整等。分类学习模块不仅要对局部态势要素进行处理形成全局态势要素,还要对历史态势要素分析处理,通过分类学习模块形成全局态势要素,进行统一存储管理。

在本文提出的态势要素获取框架中,将多 Agent 系统应用到态势要素获取,构建层次化模型。通过底层的 Agent 完成对数据的收集,位于高层的 Agent 完成对数据的分析处理。在协调管理层产生处理规则、分类规则的同时,学习形成全局态势要素,并将处理规则分发到下面各层,在各层中直接对数据进行处理,降低了数据存储空间,节省了通信能耗,减小了数据冗余属性对态势要素获取的影响。在规模化、复杂化、异构化的网络环境中,该模型是行之有效的态势要素获取架构,并为整个安全态势系统分析打下了坚实的基础。

3 安全态势要素获取方法

在层次化态势要素获取框架中,本文主要侧重于协调管理层,选用 PCA 进行属性约简,运用 PCA 消除输入数据属性之间的相关性,将高维度的特征映射到低维度形成新的特征,并通过 PCA 计算得到的特征值对 PNN 结构进行优化,减小 PNN 训练过程中的复杂度。PCA 实现简单,消耗系统资源少。选用优化的 PNN 作为基分类器,在协调管理层引入历史态势要素,运用历史态势要素对属性约简模块、增强型 PNN 进行分类训练,产生统一的属性约简规则、分类规则分发给数据采集层、汇聚分析层相应模块,减小下级各层的通信负载。具体方法模型如图 2 所示。

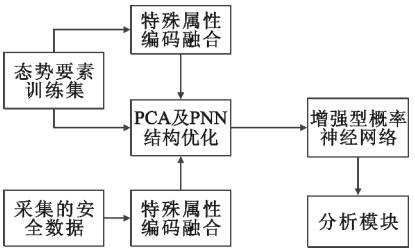

图 2 态势要素获取模型

Fig. 2 Situation element acquisition model

3.1 属性约简

PCA[11]是一种常用的特征提取方法,对数据降维并去除噪声,简化分析复杂度,发现数据模式,最大化样本方差,最大程度使新的特征互不相关。

设历史态势要素总数为 n,历史态势要素集为 $X=[x_1,x_2,\cdots,x_n]$,样本标签为 $Y=[y_1,y_2,\cdots,y_n]$,每个样本有 m 维,样本标签 y_i 分为 l 类,历史态势要素集 X 的协方差矩阵为

$$C_x=\frac{1}{n}\sum_{i=1}^{n}(x_i-E(X))(x_i-E(X))^{\mathrm{T}}=\frac{1}{\sqrt{n}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(x_i-E(X))(x_i-E(X))^{\mathrm{T}}=UU^{\mathrm{T}},\quad(1)$$

$$C_x=\mathrm{cov}(X)=\begin{bmatrix}C_{11}&C_{12}&\cdots&C_{1n}\\C_{21}&C_{22}&\cdots&C_{2n}\\\vdots&\vdots&&\vdots\\C_{n1}&C_{n2}&\cdots&C_{nn}\end{bmatrix}。\quad(2)$$

式(1)中:$E(X)=\frac{1}{n}\sum_{i=1}^{n}x_i$ 为训练集样本的平均数。设协方差矩阵 C_x 的特征值为 $\lambda_1,\lambda_2,\cdots,\lambda_n$,$C_x\in\mathbb{R}^{n\times n}$且有 $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_n$。特征向量 $U=[u_1,u_2,\cdots,u_n]$,即

$$C_xu_i=\lambda_iu_i,i=1,2,\cdots,n。\quad(3)$$

从上面的特征值排序中,选择前 p 个主元,计算其方差贡献率为

$$\eta(p)=\sum_{i=1}^{p}\lambda_i/\sum_{i=1}^{n}\lambda_i。\quad(4)$$

设 $\eta(p)\geqslant0.9$,变换矩阵 T_p 由前 p 个特征值相对的特征向量所组成

$$T_p=(t_1,t_2,\cdots,t_p)'\in\mathbb{R}^{p\times n},p<n。\quad(5)$$

变换矩阵乘以训练集数据,得到前 p 个主成分,$p_i=T_pX,i\in[1,n]$,降低了训练集的维数。

3.2 PNN 结构优化

本文运用概率乘法公式和历史态势要素集经过 PCA 后的结果对 PNN 的构造进行改进[12]。PNN 是一种 4 层前馈型并行算法的神经网络。PNN 将归

一化后的样本矩阵送入输入层,计算样本矩阵与学习矩阵之间的距离 $\| X-X_n \|$,模式层基函数以 $\| X-X_n \|$ 作为函数自变量,得到输入样本的概率,求和层将只与自身类别相关的模式层单元相连接,得到各样本属于各类的概率估计,最后竞争层求出概率最大的一类作为有效值。

在处理训练样本数很大、冗余度高的问题时,PNN 计算量大,结构复杂,系统代价高。依据概率乘法公式,本文提出了一种 PNN 结构优化方法:

$$P(ABC)=P(C/AB)P(B/A)P(A)。\tag{6}$$

当 A 和 B 独立时,式(6)可转化为

$$P(ABC)=P(C/AB)P(B)P(A)。\tag{7}$$

首先,根据 3.1 节中对训练样本集特征提取,使输入变量之间线性不相关,则可按照式(7)进行结构简化,如图 3 所示。

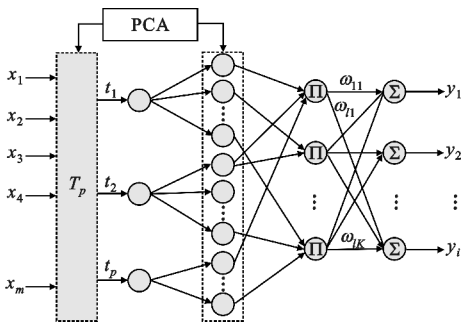

图 3　PNN 结构优化

Fig. 3 PNN structure optimization

在第二层中,每个节点代表的是对应输入样本的一种特征状态,经过 PCA 处理所得特征值调整神经元个数。特征值越大,则对应的方差越大,主成分中对判别原始训练数据有意义的信息越多。理应安排更多的模式层神经元,提高 PNN 的判别精度。假如第二层总的神经元个数为 S,则前 P 个主成分对应的第二层神经元数目为

$$q_i=\mathrm{ceil}\left[S\frac{\lambda_i}{\sum_{k=1}^{P}\lambda_k}\right],i=1,2,\cdots,P。\tag{8}$$

式中:ceil[X]为大于 X 的最小整数。第三层乘法神经元对上一层的输出求积,并与权重 ω 相乘得到此输入变量下模式层事件发生的概率。求和层通过加权求和,得出当前输入变量下所属类别的概率。

在 PNN 优化部分[12-13],其理论依据是贝叶斯最优分类决策理论,实际 PNN 输出与期望输出的误差为

$$E=\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{l}(y_j(i)-\hat{y_j}(i))^2。\tag{9}$$

式中:$\hat{y_j}(i)$ 为历史态势要素样本的原始类标签,n 为历史态势要素样本总数,l 为类别数。

模式层单元的激活函数表示为

$$R(X-X_n)=\exp\left(-\frac{1}{2\sigma^2}\| X-X_n \|^2\right)。\tag{10}$$

式中:X 为样本矩阵,X_n 为高斯函数的中心,σ 为方差。

根据 PNN 网络的结构得到网络的输出为

$$y_j=\sum_{i=1}^{K}\omega_{ij}\exp\left(-\frac{1}{2\sigma^2}\| X-X_n \|^2\right),j=1,2,\cdots,l。\tag{11}$$

需要调整的重要参数为乘法层到求和层的连接权值 ω_{ij}、模式层的中心值 X_n 和方差 σ。

连接权值 ω_{ij} 的迭代公式为

$$\frac{L_E}{L_{\omega_{ij}}}=\sum_{i=1}^{n}\sum_{j=1}^{l}(y_j(i)-\hat{y_j}(i))z_k(i),\tag{12}$$

$$\omega_{ij}=\omega_{ij}-Z\frac{L_E}{L_{\omega_{ij}}}。\tag{13}$$

式中:Z 为学习因子,$0<Z\leqslant 1$;z_k 为第三层第 k 个节点的输出。

中心值 X_n 和方差 σ 的更新公式为

$$\frac{L_E}{L_{z_k(i)}}=\sum_{j=1}^{l}(y_j(i)-\hat{y_j}(i))\omega_{ij},\tag{14}$$

$$\frac{L_{z_k(i)}}{L_{X_n}}=W_{kg}z_k(i)\frac{2}{\sigma^2}(v_g-X_n),\tag{15}$$

$$X_n=X_n-Z\frac{L_E}{L_{X_n}}=X_n-Z\sum_{i=1}^{n}\sum_{k=1}^{K}\frac{L_E}{L_{z_k(i)}}\frac{L_{z_k(i)}}{L_{X_n}},\tag{16}$$

$$\frac{L_{z_k(i)}}{L_\sigma}=W_{kg}z_k(i)\frac{2}{\sigma^3}((v_g-X_n)^2-\sigma^2),\tag{17}$$

$$\sigma=\sigma-Z\frac{L_E}{L_\sigma}=\sigma-Z\sum_{i=1}^{n}\sum_{k=1}^{K}\frac{L_E}{L_{z_k(i)}}\frac{L_{z_k(i)}}{L_\sigma}。\tag{18}$$

式中:v_g 是第二层第 g 个神经元的输入;W_{kg} 是连通示意值,代表模式层第 g 个神经元到乘法层第 k 个神经元,如果有连接 $W_{kg}=1$,否则 $W_{kg}=0$。

3.3　态势要素获取过程

本文是一个多分类问题,从传统的二分类延伸到多分类存在许多不足。在面对不止二分类问题时,传统的方法是将其变换为多个二分类问题,但算法复杂,计算和时间成本高。本文选用直接扩展到多分类的思路,在多分类问题中,由于随机评判的正确率为 $1/l$,很难发现错误率小于 1/2 的基分类器,从而导致基分类器的数量不够,组成强分类器分类效果不好[14-15]。根据杨新武等人提出的 SAMME. R

算法[16],结合上文优化后的 PNN,以此作为基分类器,每个基分类器作为一个单独的功能组件,各组件之间连续的转换集成一个强分类器,态势要素获取流程图如图 4 所示。

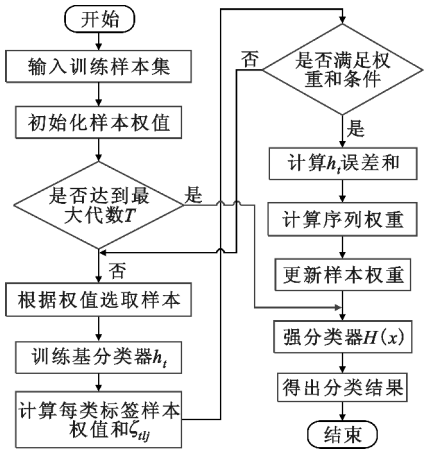

图 4 态势要素获取流程图

Fig. 4 Flow chart of situation element acquisition

本文态势要素获取算法步骤如下:

步骤 1 输入训练样本集:$(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)$,其中 $x_i\in X,y_i\in Y$,标签 y_i 类别数为 l。

步骤 2 对样本权值进行初始化:$D_1(i)=1/n$,$i=1,2,\cdots,n$。

步骤 3 设最大循环次数为 T,循环 $t=1,2,\cdots,T$:

(1)根据权值 $D_t(i)$,择取训练样本。

(2)在当前 D_t 分布下,训练优化后的 PNN 基分类器 h_t,对样本进行分类识别,$h_t:X\rightarrow Y$。

(3)循环 $l=1,2,\cdots,L$,执行下面两步。

a. 分别计算每一类时,每类样本的权值和

循环 $j=1,2,\cdots,L$

$$\zeta_{tlj}=\sum_{i=1}^{n}D_t(i)[y_i=k,h_t(x_i)=j]。\tag{19}$$

b. 判断每类中准确分类的权重和是否大于该类错误分到其他类的权重和,也就是 $\zeta_{tkj}[h_t(x_i)=j]\geqslant\forall\zeta_{tkj}[h_t(x_i)\neq j]$。如果满足条件,进入下一次循环;若不满足,返回步骤 3 重新执行。

(4)计算 h_t 的预测误差和:

$$e_t=\sum_{i=1}^{n}D_t(i)[h_t(x_i)\neq y_i]。\tag{20}$$

(5)计算序列权重:

$$a_t=\ln\left(\frac{1-e_t}{e_t}\right)+\ln(l-1)。\tag{21}$$

(6)调整下一轮样本权重:

$$D_{t+1}(i)=\frac{D_t(i)}{Z_t}\cdot\exp(-a_t\cdot h_t(x_i)\cdot y_i)。\tag{22}$$

式中:Z_t 为归一化因子,即

$$Z_t=\sum_i D_t(i)\cdot\exp(-a_t\cdot h_t(x_i)\cdot y_i)。\tag{23}$$

步骤 4 最后的强分类器为

$$H(x)=\arg\max_{y\in Y}\sum_{t=1}^{T}a_t\cdot[h_t(x_t)=y_i]。\tag{24}$$

该算法对基分类器准确率要求大于随机判断概率 $1/l$,而且每类分类正确样本权重和大于该类分到其他类的权重和。

4 实验结果与分析

本文仿真使用的数据集是 KDD CUP 99[17]测试数据集。每个样本都有相应的标签,其中包括 4 种异常类型(PROBE、DOS、U2R 和 R2L)和一种正常数据类型(Normal)。其中 Portsweep、Neptune 分别属于 PROBE 和 DOS 中的一种网络连接,且数据占有比例最大,本文选用 Portsweep、Neptune、U2R、R2L、Normal 5 种网络连接数据进行实验。每条样本由 32 维连续属性、9 维离散属性和 1 维类标签构成。

4.1 实验样本预处理

将离散文本类型数据转换成实数类型数据,实现数据类型统一。比如连接正常或错误的状态,REJ 赋值为 1,RSTO 赋值为 2 等。

在 9 维离散属性中,其中有 3 维离散特征分别是基本连接特征中的 protocol_type、service、flag,对这 3 维属性分别进行二进制编码,再融合成一个属性,转换成实数,加大了数据量纲,使得特征表现更加突出。

对连续属性类型数据归一化处理,设要处理的样本矩阵为 x_{ij},处理后的样本矩阵为 X_{b},则平均值为

$$x_i^-=\frac{1}{m}\sum_{j=1}^{m}x_{ij},\tag{25}$$

平均绝对值偏差为

$$d_i=\frac{1}{m}\sum_{j=1}^{m}|x_{ij}-x_i^-|,\tag{26}$$

标准化后的样本矩阵为

$$x_{ij}'=\frac{x_{ij}-x_j^-}{d_i}。\tag{27}$$

归一化处理之后的样本为

$$X_{\mathrm{b}}=\frac{x_{ij}'}{\sqrt{x_{i1}^{2}'+x_{i2}^{2}'+\cdots+x_{im}^{2}'}}。\tag{28}$$

实验中选用 KDD CUP 99 数据集的 10% 的训练子集和测试子集依照对应样本按照一定的比例随机抽取数据,分布情况如表 1 所示。

表 1　实验数据分布情况

Tab. 1 The distribution of test dataset

样本类别	Neptune	Portsweep	U2R	R2L	Normal	总数
训练样本 1	112	109	18	41	105	385
训练样本 2	275	124	21	78	303	801
训练样本 3	437	236	35	246	405	1 359
测试样本 4	101	104	22	32	98	357
测试样本 5	264	142	23	85	298	812
测试样本 6	468	256	42	278	612	1 656
测试样本 7	721	351	50	478	1363	2 963

4.2　实验结果

4.2.1　实验一

为了验证经过 PCA 优化后的 PNN 作为基分类器的可行性,选用训练样本 3 作为历史态势信息,测试样本 4、测试样本 5 作为测试数据,对比基本 PNN、PCA 对数据特征提取(PCA+PNN)与结构优化 PNN 对测试样本的分类精度。不同 PNN 模型对态势要素获取精度如图 5 所示。

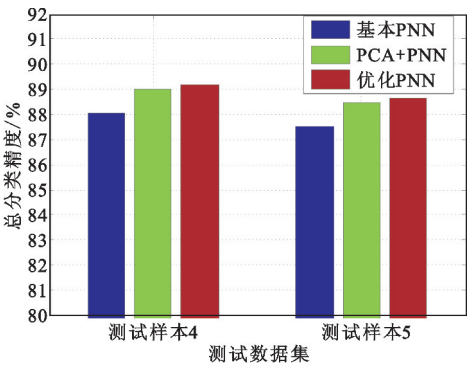

图 5　基分类器要素获取精度

Fig. 5 Base classifier element acquisition accuracy

从图中可以得出两点:一是 PCA 对数据特征提取减少 PNN 输入维数处理后总体分类精度要高于基本 PNN,表明用 PCA 作属性约简,对数据特征提取、潜在特征的挖掘是有效的,提高了要素获取精度,并且提升了分类效率、降低了数据存储开销;二是后两者的分类精度基本一致,表明优化的 PNN 能够达到 PCA 对数据特征提取减少 PNN 输入维数相当的分类效果,同时大大降低了 PNN 结构复杂度,节约了存储空间。PNN 结构复杂度可表示为 $O((n+1)d)$,其中 n 为历史态势要素样本数,d 为每个样本维数。在 n 和 d 都相当大时,需要很大的存储空间,通信消耗量也急剧上升。

4.2.2　实验二

为了分析本文设计的获取方法(简称 re-samme. r-pnn)的有效性,本文依次选用反向传播(Back Propagation,BP)神经网络、K 最近邻(k-Nearest Neighbor, KNN)、径向基函数(Radical Basis Function,RBF)神经网络、支持向量机(Support Vector Machine,SVM)和本文方案在训练样本 3 和测试样本 5 上进行实验,BP 和本文方案分别迭代 100 轮。从表 2 测试结果可以看出,本文方案的态势要素获取总准确率达到 95.93%,其各类精度均高于其他几种算法。主要原因是本方案的多分类放宽了基分类器的错误率的要求,并排除劣质基分类器,深入挖掘基分类器的分类能力。

表 2　各种不同模型的精确度

Tab. 2 The accuracy of different model　　%

方法	Neptune	Portsweep	U2R	R2L	Normal	总精确度
BP	91.19	83.25	35.60	49.52	86.81	81.89
RBF	94.21	86.37	36.14	50.26	90.53	85.32
KNN	96.13	88.24	35.27	51.35	91.47	86.57
SVM	97.31	92.93	51.56	63.67	92.70	90.06
本文方案	99.18	97.26	67.85	84.46	97.62	95.93

图 6 是以训练样本 1 作为历史态势信息,测试样本 4、5、6 作为测试集进行实验,实验结果表明对数据编码融合处理后小样本的分类精度要高于未对数据编码融合处理,这是因为对历史态势要素中的

3 维离散特征进行编码融合成一维特征,加大了特征的数据量纲,表明这 3 维离散特征对分类的贡献率很大。

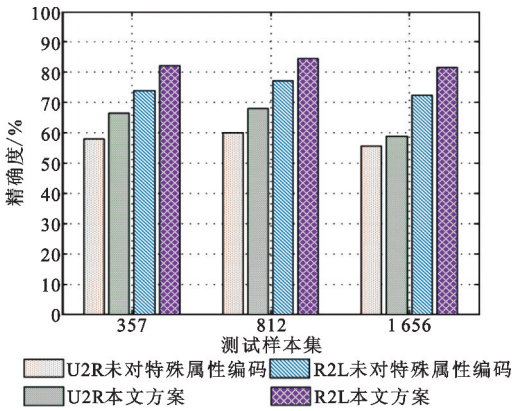

图 6 数据编码对小样本精确度的影响

Fig. 6 The influence of data coding on the small sample's accuracy

4.2.3 实验三

为了进一步验证本文方案的泛化性能,选用训练样本 1 作为历史态势信息,测试样本 4、5、6、7 作为测试集,实验结果如图 7 所示。

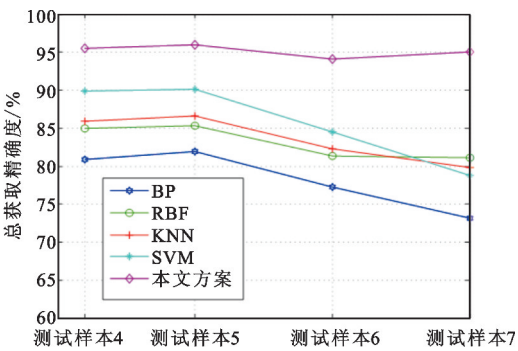

图 7 不同数据的分类精确度

Fig. 7 Classification accuracy of different data

图 7 结果表明,本文提出的方案对新样本的泛化能力总体要优于其他算法,对新样本的分类更加有效,趋于稳定,且不会出现过拟合,进一步验证了本文方案的优越性。

以上分析验证了本文设计方案的有效性和泛化能力,实验结果表明本文方案对态势要素获取分类达到了很高的精确度,为整个安全态势系统的分析提供了保障。

4.3 算法复杂度分析

按照算法的流程步骤分析时间复杂度,历史态势要素样本数为 n,样本维度为 m,标签类别数为 l,算法迭代次数为 T。本文中,l 等于 5,数量级为 1。不同算法的复杂度对比如表 3 所示。本文算法相对于 KNN、RBF 复杂度较高,但是不会造成大的影响,都在可接受范围之内。

表 3 不同算法的复杂度对比

Tab. 3 Complexity comparison among different algorithms

算法	复杂度
BP	$O(n \cdot m \cdot h_1 \cdot h_2 \cdot T)$
RBF	$O(n \cdot m)$
KNN	$O(n)$
SVM	$O(n^3)$
本文方案	$O(n \cdot m \cdot T \cdot l^2)$

注:h_1、h_2 为 BP 神经网络隐含层节点数。

5 结 论

本文针对态势要素获取困难,提出了一种增强型神经网络对态势要素进行获取;建立了基于 Agent 的分层安全态势要素获取框架,利用 PCA 对历史态势信息进行属性约简,其结果用于优化基分类器结构;对态势信息特殊属性编码融合,运用增强型多分类器实现态势要素获取。实验结果表明,本文模型对态势要素获取效果良好,提高了要素获取精度。下一步研究工作的重点是在面对大规模网络攻击时,对整个系统作出全面、实时的网络风险评估。

参考文献:

[1] 龚正虎,卓莹. 网络态势感知研究[J]. 软件学报,2010,21(7):1605-1619.
GONG Zhenghu, ZHUO Ying. Research on cyberspace situational awareness[J]. Journal of Software, 2010, 21(7):1605-1619. (in Chinese)

[2] XU Z. Demand-oriented traffic measuring method for network security situation assessment[J]. Journal of Networks, 2014, 9(4):221-224.

[3] YURCIK W. Visualizing NetFlows for security at line speed:the SIFT tool suite[C] // Proceedings of 19th USENIX Large Installation System Administration Conference. San Diego, CA:USENIX Association, 2005:169-176.

[4] SRIHARI RK. Situation awareness through concept-based information extraction[EB/OL]. (2008-07-21)[2016-04-30]. http://www. dawnbreaker. com/vas05/docs/CymfCym-Brief.

[5] BASS T. Intrusion detection systems and multisensor data fusion[J]. Communications of the ACM, 2000, 43(4):99-105.

[6] WANG H,LIANG Y,YE H. An extraction method of situational factors for network security situational awareness [C]// Proceedings of 2008 International Conference on Internet Computing in Science and Engineering. Washington,DC:IEEE,2008:317-320.

[7] 郭文忠,林宗明,陈国龙. 基于粒子群优化的网络安全态势要素获取[J]. 厦门大学学报(自然科学版),2009,48(2):202-206.
GUO Wenzhong,LIN Zongming,CHEN Guolong. Network security situation element extraction based on particle swarm optimization[J]. Journal of Xiamen University(Natural Science Edition),2009,48(2):202-206. (in Chinese)

[8] LI D,LIU Z. Situation element extraction of network security based on logistic regression and improved particle swarm optimization[C] //Proceedings of 2012 International Conference on Natural Computation. Shenyang:IEEE,2013:569-573.

[9] LEETON U,KULWORAWANICHPONG T. Multi-agent based optimal power flow solution[C]//Proceedings of 2012 Power and Energy Engineering ConferencE. Shanghai:IEEE,2012:1-4.

[10] MA L,TSAI J. Formal modeling and analysis of a secure mobile-agent system[J]. IEEE Transactions on Systems Man and Cybernetics-Part A: Systems and Humans,2008,38(1):180-196.

[11] 冯祥,陈良彬. 基于主成分分析和独立成分分析的调制分类算法[J]. 电讯技术,2013,53(7):864-867.
FENG Xiang,CHEN Liangbin. Modulation classification algorithm based on PCA and ICA[J]. Telecommunication Engineering,2013,53(7):864-867. (in Chinese)

[12] 邢杰,萧德云. 基于 PCA 的概率神经网络结构优化[J]. 清华大学学报(自然科学版),2008,48(1):141-144.
XING Jie,XIAO Deyun. PCA-based probability neural network structure optimization[J]. Journal of Tsinghua University(Science and Technology),2008,48(1):141-144. (in Chinese)

[13] 李春芳,刘连忠,陆震. 基于数据场的概率神经网络算法[J]. 电子学报,2011,39(8):1739-1745.
LI Chunfang,LIU Lianzhong,LU Zhen. Probabilistic neural network based on data field[J]. Acta Electronica Sinica,2011,39(8):1739-1745. (in Chinese)

[14] ZHU J,ZOU H,ROSSET S,et al. Multi-class adaboost [J]. Statistics and Its Interface,2009,2(3):349-360.

[15] APPEL R,FUCHS T,DOLLÁR P,et al. Quickly boosting decision trees-pruning underachieving features early[C]// Proceedings of 2013 International Conference on Machine Learning. Atlanta,GA:ICML,2013:594-602.

[16] 杨新武,马壮,袁顺. 基于弱分类器调整的多分类 Adaboost 算法[J]. 电子与信息学报,2016,38(2):373-380.
YANG Xinwu,MA Zhuang,YUAN Shun. Multi-class adaboost algorithm based on the adjusted weak classifier[J]. Journal of Electronics and Information Technology,2016,38(2):373-380. (in Chinese)

[17] KDD CUP 1999 data[EB/OL]. (2007-06-26)[2016-04-30]. http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html.

作者简介:

李方伟(1960—),男,重庆人,教授、博士生导师,主要研究方向为移动通信理论与技术、信息安全技术;
Email:lfw@cqupt.edu.cn

王　森(1990—),男,重庆人,硕士研究生,主要研究方向为网络安全态势感知;
Email:senma1990@163.com

朱　江(1977—),男,湖北人,2009 年于电子科技大学获博士学位,现为副教授,主要研究方向为认知无线电;

张海波(1979—),男,重庆人,博士,副教授,主要研究方向为无线资源优化。