基于 Kohonen 神经网络算法的网络入侵聚类算法的测试研究

麻书钦
(广东技术师范学院,广东 广州 510665)

摘 要:提出一种基于 Kohonen 网络的网络入侵聚类研究的方法,在阐述基本理论、原理和算法步骤基础上,利用 Matlab 软件平台对提出的网络入侵算法进行测试研究,并同其他方法进行仿真对比,发现 Kohonen 神经网络算法的网络入侵聚类在训练准确率、测试准确率和运行时间 3 个方面都优于 PNN 算法,其准确率可以达到 98.1%。

关键词:Kohonen 神经网络;网络入侵;Matlab 软件;聚类算法
中图分类号:TP393;TP391.9;TP183;TP393.08 **文献标志码**:A **文章编号**:1674-5124(2013)04-0113-04

Research on network intrusion clustering based on Kohonen neural network algorithm

MA Shu-qin
(Guangdong Polytechnic Normal University, Guangzhou 510665, China)

Abstract: This paper presents a method of clustering of network intrusion based on Kohonen network. Firstly, the basic theory, principle and algorithm steps are introduced. Then, matlab software platform was used for testing the proposed network intrusion algorithm. Finally, this algorithm was compared with other methods though simulation tests. Experimental results show the Kohonen neural network clustering algorithm is better than PNN algorithm in three aspects, i.e., training accuracy, testing accuracy and operation time, its accuracy rate can reach 98.1%.
Key words: Kohonen neural network; network intrusion; matlab; clustering algorithm

0 引 言

随着信息技术和网络技术的快速发展,网络非法入侵也随之大幅增长,对网络安全性提出了强大挑战,因此维护网络安全显得尤为重要。对网络入侵聚类的研究[1],有助于应对网络入侵,在此基础上提出合适的防控措施。根据 Kohonen 网络的自动聚类的优点,本文提出一种基于 Kohonen 网络的网络入侵聚类研究的方法。首先阐述基本理论、原理和算法步骤,然后利用 Matlab 软件平台对提出的网络入侵算法进行测试研究,并同其他方法进行对比,从而验证出 Kohonen 神经网络算法进行网络入侵聚类的优越性和准确性。

收稿日期:2013-02-25;**收到修改稿日期**:2013-04-18
基金项目:全国教育科学"十二五"规划 2012 年度教育部重点课题(DCA120190)
作者简介:麻书钦(1975-),男,广西横县人,讲师,硕士,研究方向为计算机网络管理、校园网络维护、虚拟化云平台。

1 Kohonen 网络结构

Kohonen 网络是一种自组织无监督学习网络,可以识别环境特征,同时实现自动聚类。该网络由芬兰学者 Kohonen 提出[2-3],调整网络权值主要通过自组织特征映射完成,从而实现神经网络收敛。

Kohonen 网络由两层前馈神经网络组成,包括输入层和输出层。输入层和输出层之间通过神经元进行双向连接,将输入在输出层映射成二维离散图像。Kohonen 网络拓扑结构如图 1。

输入层:用以表现网络的输入变数,即训练范例的输入向量,或称特征向量,其处理单元数目依问题而定,每一个处理单元代表着输入向量的每一个元素,亦即该输入资料所拥有的特征。

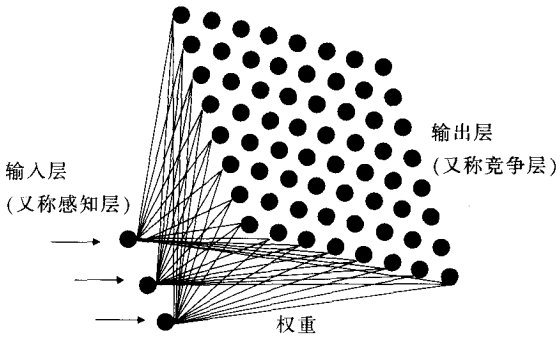

图1 Kohonen神经网络结构图

输出层：用以表现网络的输出变数及训练范例的聚类，其处理单元数目依问题而定。其结构本身有“网络拓扑”以及“邻近区域”的概念。

网络连结：每个输出层处理单元与输入层处理单元相连结的权数所构成的向量，表示一个输入特征值向量对应训练范例聚类的标量。当 Kohonen 网络学习完毕后，靠近输出处理单元的神经元具有相似的连结权数。

2 Kohonen 神经网络的原理

Kohonen 网络的基本原理是计算输入的特征量映射至输出层每一处理单元的欧几里得距离(euclidean distance)，而具有最小距离值的处理单元就是优胜单元并且将会调整它的连接权值，使其能够更接近原始的输入向量，而且此处理单元的邻近区域也会调整本身的连接权值，使自己与输入向量间的欧几里得距离能够减少，其算法步骤[4]如下：

输入：训练样本和测试样本；

输出：训练后的权系数矩阵和测试样本所属的类及归属程度；

(1)粗调整学习阶段

1)网络权值初始化 w_{ij}，新向量的输入

$$\boldsymbol{X}=[X_1(t),X_2(t),\cdots,X_m(t)]^T \quad (1)$$

式中：$X_i(t)$——t 时刻样本的第 i 维分量 $(i=1,2,\cdots,m)$，总的学习次数为 $T=T_1+T_2$。

2)样本矢量与权值之间距离的计算

$$d_j=\|X-w_j\|=\sqrt{\sum_{i=1}^{m}(x_i-w_{ij}(t))^2}$$

$$(j=1,2,\cdots,n) \quad (2)$$

3)求最小距离，找出最匹配输入样本矢量的竞争层节点 c，即：

$$d_c=\min d_j \quad (j=1,2,\cdots,n) \quad (3)$$

4)调整权系数，粗调整阶段

$T_1=100$，$\eta_0=0.4$，$\sigma_0=3m/2$

其中：m——竞争层神经元的个数。

按照步骤 2)计算邻域函数值，权值可根据式(4)进行调整：

$$w_{ij}(t+1)=w_{ij}(t)+\eta(t)N_{ic}(t)(x-w_{ij}(t)) \quad (4)$$

5)学习率和邻域宽度按照步骤 1)和步骤 3)进行递减。

6)返回步骤 3)，所有学习样本调整一遍。

7)$t=t+1$；直至 $t>T_1$。

(2)精细调整学习阶段

8)精细调整 $T_2=500$ 阶段，$\eta_0=0.04$，$\sigma_0=1$，并重复步骤 2)~6)，只是邻域宽度和学习率按照式(5)递减：

$$\eta(t)=\eta_0(1-t/T_2)$$
$$\sigma(t+1)=\sigma_0(1-t/T_2) \quad (5)$$

9)将另一组样本矢量作为网络输入，返回第 3)步，直到样本输入结束。

10)$t=t+1$；当 $t>T_2$ 时，学习阶段结束。

11)所有输出神经元的连接权系数的存储和输出。

(3)应用阶段

12) 输入样本和连接输出神经元的权系数向量的归一化，以保证输入样本和输出神经元的欧氏距离在[0,2]之间，如式(6)所示：

$$x'_{ij}=\frac{x_{ij}}{\sqrt{x_{1j}^2+x_{2j}^2+\cdots+x_{mj}^2}}$$
$$w'_{ij}=\frac{w_{ij}}{\sqrt{w_{1j}^2+w_{2j}^2+\cdots+w_{mj}^2}} \quad (6)$$

13)读取一个新的网络输入样本。

14)根据已计算出来的欧氏距离 d_j 和上文定义的隶属函数，计算输入样本到各输出神经元的隶属度为

$$\mu(C_j(X))=\mu(d_j(X),0,\sqrt{0.2})=e^{-\frac{d(X)^2}{0.4}} \quad (7)$$

15)输出分类：设定阈值，根据高斯函数自身性质，本文将阀值设为 0.6，即：当 $\mu(C_j(X))>0.6$ 输出神经元即为该输入样本所属的类，输出该类及输入样本在该类中的隶属度。

16)重复 13)步，直到没有新的输入样本。

根据算法步骤可知其算法流程图如图 2 所示。

3 Kohonen 神经网络网络入侵算法的构建

3.1 网络入侵

网络入侵是指通过非法手段试图破坏计算机和网络系统资源完整性、机密性和可用性的行为。对网络入侵进行聚类分析研究，有助于发现网络入侵的种类，在此基础上为制定防网络入侵的措施和方案，提供决策依据。

3.2 模型建立

根据网络入侵的特点，Kohonen 神经网络网络入侵攻击聚类算法流程如图 3 所示。

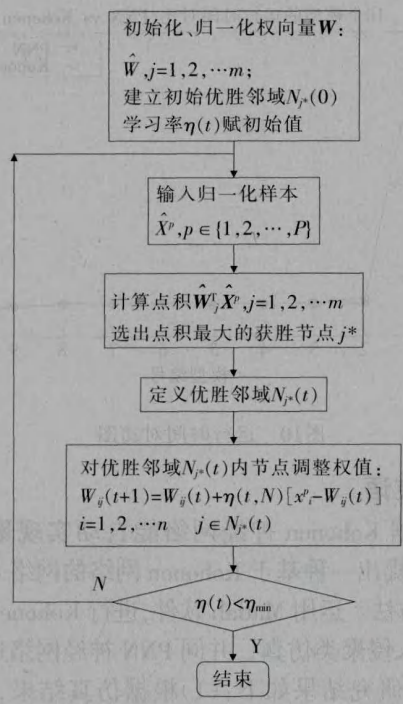

图2　Kohonen网络算法的程序框图

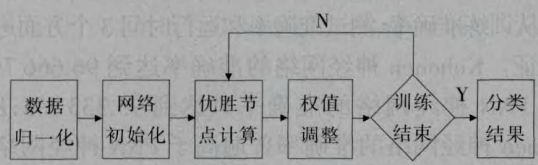

图3　算法流程图

3.3　算法的实现与测试

根据某具体的项目,现有 5000 组网络入侵数据,数据中有 5 类入侵方式,分别为 I 类、II 类、III 类、IV 类、V 类。用训练好的 Kohonen 神经网络测试样本数据,测试数据有 500 组,运用 Matlab 进行相应的仿真,仿真结果如图 4~图 7 所示。

图 4 中的神经元编号方式是从左到右,从下到上,神经元编号不断增加,左下角的神经元为 1 号,右上角神经元为 16 号,中间的数字代表神经元的获胜次数。图 5 表示网络权值的分布。

图 6 表示邻近神经元的距离分布图,相邻神经元间填充的颜色表示两个邻近神经元的距离远近,颜色越深,越接近黑色,代表距离越远,反之越近。

从图 7 可以看出,绝大多数测试结果同预期结果一致,预测结果的分类准确率达到 98.1%,效果很好。

4　不同聚类算法的测试对比

为了进一步验证 Kohonen 神经网络网络入侵聚类方法的优越性和准确性,将其同 PNN 神经网络算法进行对比,主要从训练准确率、测试准确率和运行

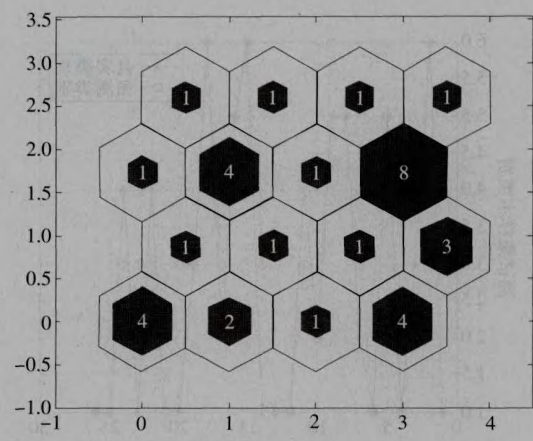

图4　获胜神经元统计图

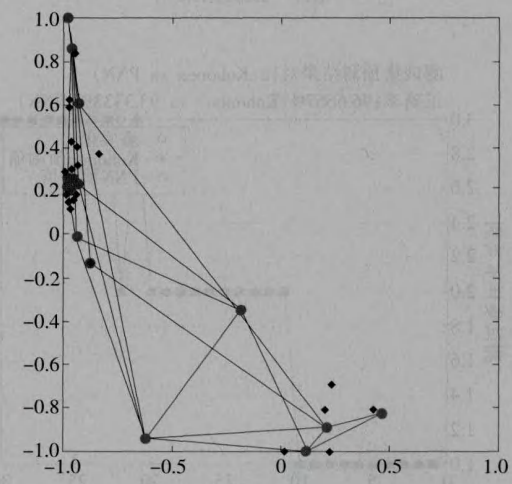

图5　网络权值分布

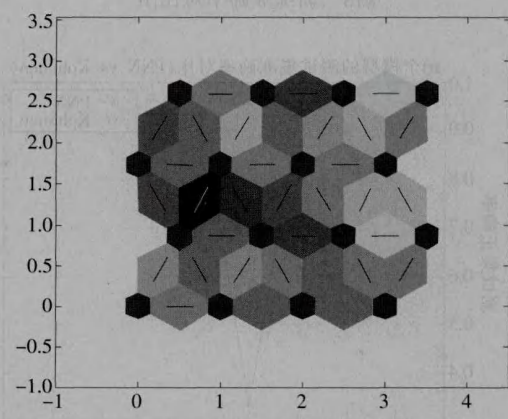

图6　邻近神经元距离分布图

时间 3 个方面[5-7]进行验证,仿真结果如图 8~图 10 所示。

从图 8 可以看出,Kohonen 神经网络的准确率有 96.6667%,而 PNN 神经网络[8-9]的准确率只达到 93.3333%。从图 9 可以看出,Kohonen 神经网络的准

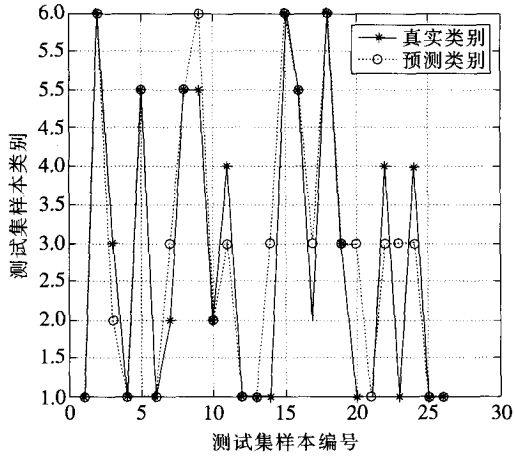

图7 预测结果

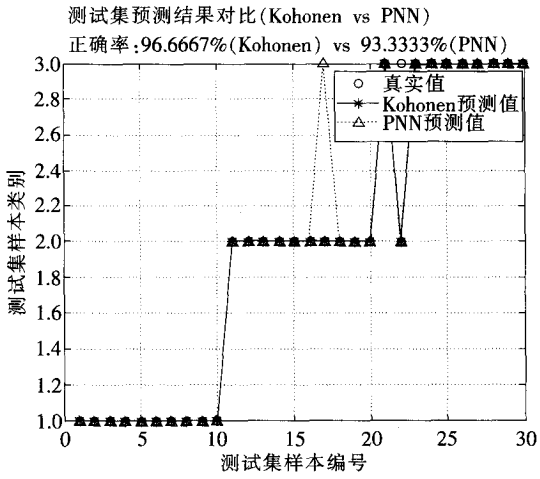

图8 训练准确率对比图

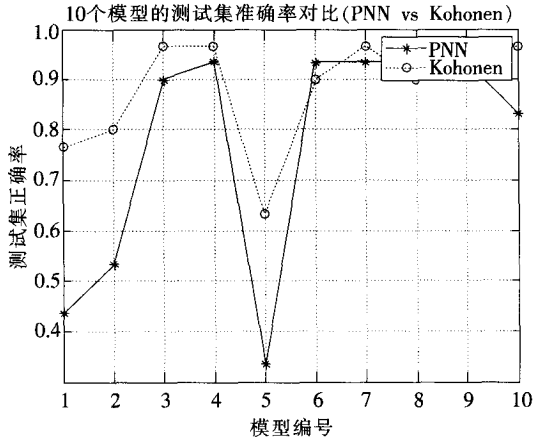

图9 测试准确率对比图

确率普遍高于 PNN 神经网络的准确率。从图 10 可以看出,Kohonen 神经网络的运行时间也优于 PNN 神经网络。

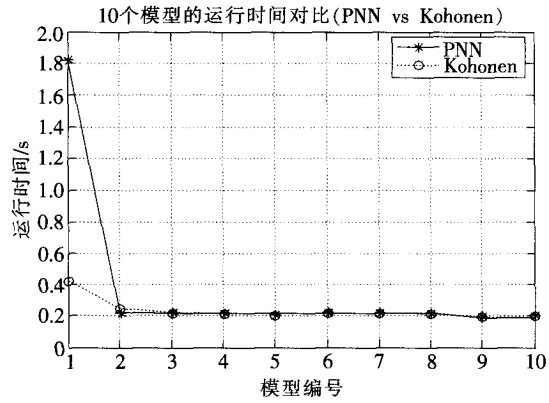

图10 运行时间对比图

5 结束语

根据 Kohonen 神经网络能自动实现聚类的优点,本文提出一种基于 Kohonen 网络的网络入侵聚类研究的方法。运用 Matlab 软件,进行 Kohonen 神经网络网络入侵聚类仿真,并同 PNN 神经网络进行了对比,主要研究结果如下:(1)根据仿真结果,Kohonen 神经网络网络入侵聚类结果的准确率达到 98.1%,效果很好。(2)将其同 PNN 神经网络算法进行对比,主要从训练准确率、测试准确率和运行时间 3 个方面进行验证。Kohonen 神经网络的准确率达到 96.666 7%,而 PNN 神经网络的准确率只达到 93.333 3%;Kohonen 神经网络的准确率普遍高于 PNN 神经网络的准确率;Kohonen 神经网络的运行时间也优于 PNN 神经网络。

参考文献

[1] 樊玫. 基于 Kohonen 神经网络的用户访问模型挖掘模式的研究[D]. 南昌:南昌大学,2007.

[2] 刘纯平. 基于 Kohonen 神经网络聚类方法在遥感分类中的比较[J]. 计算机仿真,2006,26(7):1744-1746.

[3] 范作民,白杰,阎国华. Kohonen 神经网络在发动机故障诊断中的应用[J]. 航空动力学报,2000,15(1):89-92.

[4] 莫礼平. 基于 Kohonen 神经网络的故障诊断方法[J]. 成都大学学报,2007,(1):47-51.

[5] Agrawal R, Srikant R. Mining sequential pattems[C]// International Conference on Data Engineefing.Taipei,Taiwan:ICDE,1995:3-14.

[6] Francesco B, Fosca G, Giuseppe M, et al. Data Mining for intelligent web caching[C]//International Conference on Information Technology:Coding and Computing,2001.

[7] 吴柯,方强,张俊玲,等. 基于改进 Kohonen 神经网络的遥感影像分类[J]. 测绘信息与工程,2007,32(2):47-49.

[8] 李宗福,邓琼波,李桓. Kohonen SOFM 神经网络及其演化研究[J]. 计算机工程与设计,2004,25(10):1729-1730.

[9] 曲义飞. 基于 Web 使用挖掘的用户消费模式发现研究[D]. 大连:大连理工大学,2006.