ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 32,No. 3,May 2017,pp. 479—488
DOI:10. 16337/j. 1004-9037. 2017. 03. 006

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:+86-025-84892742

机器学习在网络入侵检测中的应用

朱 琨 张 琪

(南京航空航天大学计算机科学与技术学院,南京,211106)

摘 要: 随着网络的快速发展,网络安全成为计算机网络中一个重要的研究方向。网络攻击日益频繁,传统的安全防护产品存在漏洞,入侵检测作为信息安全的重要防护手段弥补了防火墙的不足,提供了有效的网络入侵检测措施,保护网络安全。然而传统的入侵检测系统存在许多问题,基于机器学习的入侵检测方法实现了对网络攻击的智能检测,提高了入侵检测的效率,降低了漏报率和误报率。本文首先简要介绍机器学习的部分算法,然后对机器学习算法在网络入侵检测中的应用进行深入的分析,比较各个算法在入侵检测应用中的优势和缺点,最后总结了机器学习的应用前景,为获得性能良好的网络入侵检测和防御系统奠定基础。

关键词: 机器学习;网络入侵检测;决策树;神经网络;支持向量机

中图分类号: TP393 **文献标志码:** A

Application of Machine Learning in Network Intrusion Detection

Zhu Kun, Zhang Qi

(College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, 211106, China)

Abstract: With the development of network, network security becomes the key course of computer research. Hacker attacks become more and more frequent. The traditional security products have loopholes. Intrusion detection, as an important means of information security, makes up for the shortcomings of the firewall, provides an effective network intrusion detection measures and protects the network security. However, there are a lot of problems in traditional network intrusion detection. Methods based on machine can detect network intrusion automatically, improve the efficiency of intrusion detection, and reduce the false negative rate and false alarm rate. Here, we first introduce some machine learning algorithms briefly, and then analyze the application of machine learning algorithm in network intrusion detection. Moreover ,we compare the advantages and disadvantages of each algorithm applied in intrusion detection. Finally we summarize the application prospect of machine learning to lay the foundation for the network intrusion detection and prevention system with good performance.

Key words: machine learning; network intrusion detection; decision tree; neural network; support vector machine

---

基金项目: 江苏省自然科学基金(BK20160812)资助项目。
收稿日期: 2016-05-09;**修订日期:** 2016-10-14

引 言

当今社会网络成为人类交流的常用工具,因此对网络的安全要求很严格。计算机网络实现了各类金融、资源管理以及虚拟社区的应用,加快了社会信息化的进程。然而网络的对外开放性为网络攻击者提供了破坏的机会。网络入侵者通常尝试从网络中获得信息、修改信息或者造成系统瘫痪。例如拒绝服务,攻击者试图使主机无法获得资源,从而不能正常工作。网络病毒通过网络感染其他主机。面对严峻的网络环境,网络入侵检测已经成为计算机网络一个重要的研究方向。在一个网络或一个系统中,任何形式未经授权的或未经批准的活动被称为入侵[1]。1980 年,美国教授 James 在撰写《Computer Security Threaten Monitor》论文时对网络威胁进行了分析,并第 1 次提出了入侵检测[2]。网络入侵检测系统帮助发现和识别未经授权的使用、复制、修改和破坏[3]。从网络安全的角度来分析,入侵检测作为网络防御的第 2 道门,具有十分重要的作用。网络入侵检测方法分为异常检测和误用检测,异常检测首先构建一个正常模型,不符合该模型的访问均被定义为入侵。相反误用检测根据不可接受的行为建立一个入侵模型,符合该模型的访问均为入侵。然而面对日益复杂的网络环境,传统的入侵检测系统已经难以招架,如系统占用资源过多,对未知的网络监测能力不佳;异常检测的误报率高,误用检测的漏报率高;对异常的数据分析不足,需要人工干预等缺陷日益突出。入侵检测系统必须根据外界的攻击进行自我学习,机器学习算法致力于使智能体模拟人类的行为,随着经验的提升来提高自身性能,学习新知识,获得新技能。网络入侵检测系统引入机器学习的方法,更具智能性,提高了检测效率,降低了检测的漏报率和误报率。机器学习最早在 20 世纪 50 年代后期以人工智能技术的概念提出[4],网络入侵检测系统中引入机器学习的方法,主要是将入侵检测问题转化为模式识别和分类的问题来处理,采用机器学习的算法对网络中的正常行为和异常行为进行分类。整个处理过程主要分为:获取特征信息并进行特征分析;学习特征提取后的数据,进行深入分析;判断数据类型。基于机器学习的方法(神经网络[5]、支持向量机[6]和决策树[7]等)在提高入侵检测系统性能方面已经取得了不俗的成就。目前机器学习在网络入侵检测中的应用已经受到了各界的广泛关注,国内外很多著名学者和研究机构都投入到了相关技术的研究中,在基本理论、关键技术和架构等方面取得了代表性成果。

1 基于机器学习的网络入侵检测技术

随着网络的快速发展,网络环境趋于多样化、复杂化,传统的网络入侵检测技术已经难以抵挡各式各样的网络攻击。大数据时代的到来迫使网络入侵系统采用更加高效的算法。机器学习中大部分的算法主要用于解决分类问题,能够将网络行为进行不同的划分,并且机器学习算法善于使智能体模拟人类的行为,通过自身的学习,提高自身的性能,十分适用于当前多样复杂的网络环境。网络入侵检测系统中引入机器学习的算法,使得系统更加智能化、高效化,提高了入侵检测的正确率,保护了网络的安全。

1.1 决策树

决策树是常用的分类器之一。算法通过训练集构建决策树,利用决策树对测试集进行分类。分类方法通过学习树迭代输入的数据来预测数据的标签[8]。决策树的构造过程不依赖领域知识,关键步骤是分裂属性,即在某个节点处按照某一个特征属性的不同划分不同类的属性。该算法简单易理解,能够在较短的时间内处理大型数据源。假使将网络入侵检测系统转化为判断网络行为是否为网络入侵行为或者属于哪种网络入侵行为的过程,那么基于决策树的网络入侵检测系统就可以利用该算法构建决策树,并且进行入侵预测。决策树在实时入侵检测中应用效果良好,模型易于构建和解释,并且提高了网络入侵检测效率。2009 年,文献[9]提出了一种实时检测方法,采用数据包嗅探器每两秒嗅探网络数据包,并预处理 12 个特征,用决策树算法对网络数据进行分类。输出结果可以分为拒绝服务、彻底调查和

正常与结果显示,该算法具有 97.5%的检测率。但是该模型并不能检测未知攻击。2010 年,Abbes 等应用决策树和协议分析进行有效的入侵检测[10],为每个应用层协议构建一个自适应决策树,异常检测将数据分为两类:良性和异常。2012 年,Sinapiromsaran 等提出了多属性框架决策树[11],从最远的一对中选择一个核心向量,并且将数据分为左区域、右区域和中部区域共 3 个区域来帮助将恶意访问进行分类。在基于决策树的网络入侵检测研究中,大部分研究人员采用 KDD Cup 99 数据集,该数据集不能显示当前的网络情况。文献[12]采用了 Kyoto 2006^{+} 数据集,其中每个样本被标记为正常、已知攻击和未知攻击。作者利用决策树(J48)算法对网络包进行分类,并且对 134 665 个样本进行训练和测试。决策树的构造过程中最重要的是分裂属性。元组本身具有多个属性,在构建的过程中需要判断对何种属性进行分类,属性选择度量的方式有信息增益、增益率等。该论文采用信息增益作为属性选择度量。属性 A 的信息增益表达式为

$$\text{gain}=\text{inf}(T)-\sum_{i=1}^{S}\frac{|T_i|}{T}\times\text{inf}(T_i) \tag{1}$$

式中:T 为全部数据集,T_i 为 T 上为属性 A 组成的不同值的自集合,$\text{inf}(T)$ 为熵函数,计算公式为

$$\text{inf}(T)=-\sum_{j=1}^{\text{Nclass}}\frac{\text{freq}(C_j,T)}{|T|}\times\log_2\left(\frac{\text{freq}(C_j,T)}{|T|}\right) \tag{2}$$

本节主要描述了决策树的算法以及研究人员利用决策树在网络入侵检测中的一些应用,重点介绍了构建决策树的重要步骤:分裂属性,并给出了其中一种属性选择度量的方式:信息增益的计算公式。决策树算法中,分类的过程顺着其中的一条支线向下,并且属性选择度量保证了创建决策树的最优化,与传统的入侵检测方法相比,基于决策树的入侵检测系统分类效率得到了显著的提高。

1.2　神经网络

神经网络的灵感来源于人脑,由相互连接的人工神经元组成,并且能够对它们的输入进行一定的计算[13]。神经网络由许多的神经元相互连接形成。每个神经元代表一种特定的输出函数,称为激励函数。每两个节点之间有一个权重,开始时,所有的权重初始为任意值,而后根据一系列的输入输出来调节权重。图 1 表示典型的 3 层神经网络模型,从图 1 中可知,网络共包含 3 层。输入层的输出和隐藏层的输出分别作为隐藏层每一个神经细胞的输入和输出层的输入。在实际运用中,隐藏层有可能不止一层,但多数情况下,一层已足够解决问题。神经网络实现了高速并行计算,有利于解决传统网络入侵检测系统中中央节点负荷过大造成单节点失效的问题;具有较强的联想能力和自组织能力,能够预测网络攻击,有利于弥补传统网络入侵检测系统缺乏主动防御能力的问题;能够处理大规模的数据,具有分布式存储、弹性拓扑等特点,有利于优化传统网络入侵检测系统,为现代网络新环境(如云计算)提供安全保障。文献[14]提出了一个基于神经网络的入侵检测系统,该系统基于非监督的神经网络,目的是对网络进行智能实时入侵检测。整个系统的框架见图 2。系统的第 1 部分捕捉和预处理实时网络流量数据,提取数据特征并且转换成二进制或标准化形式;转化后的数据送入基于神经网络的检测系统,该系统使用自适应共振理论(Adaptive resonance theoty,ART),自组织映射(Self-organizing map,SOM)[15]和神经网络;最后输出结果写入日志,如果检测到异常就发出警报。神经网络的训练过程如图 3 所示。起初神经网络自动学习并且根据输入数据的相似性进行分类。聚类完成后,系统确定每个群集的神经元,为每个集群分配 1 个来自于数据包标签的名字。具有相同名称的群集组成 1 个单元。上述过程构建了一幅聚类图。在这幅图中,被聚集在一起的单元可以表示正常、已知的攻击

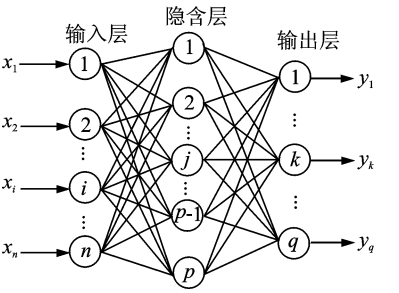

图 1　典型 3 层神经网络模型

Fig. 1　Typical three-layer neural network model

或者可能性的新攻击。较早期间，Aljurayban 和 Emam[16] 使用分层异常检测框架来有效保护云网络环境，而这个框架通过人工神经网络创建一个用于检测的数据挖掘知识库。有效的检测和以较少的流量分析达到增长的吞吐量。分层异常检测框架能够处理大量数据流，并且能够在更加苛刻的环境中保持云网络和服务的有效运行。Barollid 等[17]将神经网络作为 Tor 网络上异常检测系统解决方案研究入侵检测系统的使用，使用 Tor 服务器和客户端以及反向传播神经网络在 Tor 网络上模拟交易并且进行数据采集。该系统提出一个从 Wireshark 捕捉的数据来训练神经网络，然后将服务器数据和客户端数据进行比较，差异被认为是入侵。测试的结果在测试环境中进行评估时证明准确性好。在本小节中，主要描述了神经网络的算法，以及研究人员利用神经网络在网络入侵检测中的一些应用，举例介绍了神经网络在入侵检测过程中的作用和优点。在具有虚拟化、分布式和超大规模等特点的云环境下，基于神经网络的网络入侵检测系统方面的研究具有较强的学术和实践价值。

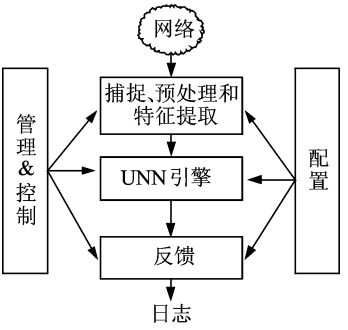

图 2 系统架构

Fig. 2 System architecture

1.3 支持向量机

支持向量机是一个分类器，在特征空间中寻找一个超平面，这个超平面分割了两个类并且与每个类的边缘距离最大[18]。通俗来说支持向量机是一种二分类模型，学习策略便是间隔最大化，最终转化为凸二次规划问题的求解。其分类问题大致有 3 种，包括线性可分问题、近似线性可分问题和线性不可分问题。图 4 线性分类的例子，虚线上的点称为支持向量。线性不可分是指一个数据集不能通过一个线性分类器实现分类。在实际的环境中，经常会遇到线性不可分的情况，例如：人脸识别、文本文档等。常用的解决方法是使用核函数，其本质思想是将样例特征映射到高维空间中。映射后的维度可能太高而导致计算复杂化，核函数的作用是在转化之前在低维上进行计算，而将实质上的分类效果表现在高维上，降低了计算的复杂性。与 1.2 节介绍的神经网络不同，支持向量机学习问题可以表示为凸优化问题，利用已知的算法求解目标函数的全局最小值，神经网络采用基于贪心学习的策略，求解得到局部最优解。支持向量机原理如图 4 所示。在网络入侵检测中引入支持向量机，在先验知识不足的情况下，支持向量机依然有很好的分类正确率[19]，并且具有较强的推广能力。根据过去的数据或经验得出规律，用学习到的规律对未知的网络行为作出正确的预测，这种能力也称为推广能力。2009 年，Li 等为了提高无线自组织网络中异常探测的准确性，提出了一种新的模糊支持向量机网络（Fussy discrimination support vector machines，SVMFN）[20]。实验的结果显示 SVMFN 要比传统的方法准确率高，并且更加适应工程应用的环境。Li 等在此基础上考虑实用性，结合基于二叉树的多类支持向量机[21]提高了异常检测的准确性和训练时间。Zhang 等同样也结合多类支持向量机应用于合作网络入侵检测系统[22]。为了提高支持向量机的分类效率，可采取减少故障属性、降低分类空间维度的措施。Guo Jiangwei[23] 和 Liu Zhiguo[24] 等在不同的论文中将粗糙集和支持向量机进行结合，论文中采用粗糙集理论降低了故障的属性，从而减少了分类空间的维度。文献[24]中的实验结果表明结合了粗糙集理论的支持向量机分类效率要比单一的支持向量机算法分类效率高很多。此外，Lin Nan

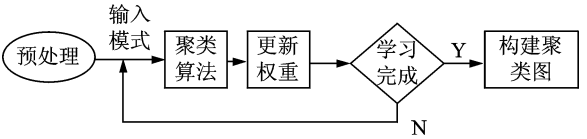

图 3 训练过程

Fig. 3 Training process

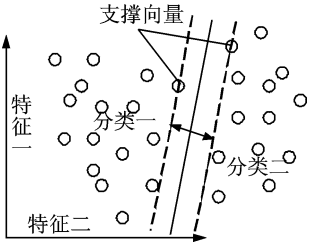

图 4 支持向量机

Fig. 4 Support vector machine

等提出了一种基于小波变换的支持向量机集成算法[25],使用小波变换将原始数据集的冗余属性删除,在简化的数据集中进行训练。通过将支持向量机与其他理论或技术相结合,同样可以提高支持向量机的分类效率,从而提高网络入侵检测系统的性能。2010 年,文献[26]提出了一种最小二乘支持向量机,并将其应用于网络入侵检测系统之中,最小二乘支持向量机可以解决凸二次规划问题。Wang 等将遗传算法与支持向量机相结合[27],在支持向量机算法中,不恰当的训练参数会导致过拟合或欠拟合,遗传算法用来选择适当的支持向量机训练参数。2015 年有学者提出一种新的网络入侵检测方法,将混合蚁群算法和支持向量机结合[28],从而保证网络入侵检测的高精准性。

1.4　贝叶斯分类

贝叶斯分类是一类以贝叶斯公式为基础的分类方法的总称,贝叶斯定理用于解决已知 $P(A|B)$,求取 $P(B|A)$的情况。$P(A|B)$表示在事件 B 已经发生的情况下,事件 A 发生的概率。在机器学习中,最常见的是分类问题,在给定数据的前提下,判断该数据属于某个分类的概率。显而易见,贝叶斯定理适用于解决该类问题。贝叶斯分类算法中最简单的是朴素贝叶斯分类器,其基本思想为:对于待分类项,计算该项属于各个分类的概率,判断代分类项属于概率最大的分类。朴素贝叶斯简单易理解,但有局限性,其特征属性必须条件独立或基本独立,在实际应用中,这样的前提条件很难得到。贝叶斯分类中较常用的为贝叶斯网络,贝叶斯网络的定义包含一个有向无环图和条件概率表集合,并且若前驱节点值确定后,后续的节点独立于所有前驱节点。由此,得出任意随机变量组合的联合条件概率分布为

$$P(x_1,x_2,\cdots,x_n)=\prod_{i=1}^{n}P(x_i\mid \mathrm{Parents}(x_i)) \tag{3}$$

式中:Parents 表示 x_i 的直接前驱节点的联合。贝叶斯算法是机器学习分类算法中最常用的一种,具有很高的分类效率和很强的扩展性,将贝叶斯算法运用于入侵检测系统中可以提高入侵检测技术的性能和加强检测模型的稳定性。如图 5 所示,基于贝叶斯分类算法的网络入侵模型主要分为 3 个模块:预处理模块、检测模块[29]和响应模块。这种入侵检测技术具有高效处理大数据的能力,并且通过机器学习的特性不断地更新和完善样本集从而提高攻击匹配率,对未知特征的攻击类型有较高的检测率。将贝叶斯算法运用到网络入侵检测技术中构建的模型结构简单、分类速度快、准确度高且能够自我完善。2012 年,Carol J Fung 提出了一个协同入侵检测网络[30],该网络允许分布式的入侵检测系统协作并且分享各自关于入侵的认知和观点,从而增强入侵检测的整体精确性和检测新入侵分类的的能力。在这个过程中,每个入侵检测系统通过贝叶斯学习方法评估邻居检测系统的检测准确性和错误率,并且将这些结果进行聚类。Guan Keqing 等构建了一个基于贝叶斯判别方法的入侵检测模型[31],将入侵检测问题转化为判别分类问题。基于贝叶斯算法的入侵检测运用于不同的网络环境当中,Chirag N Modi 等在云环境中设计和集成了基于贝叶斯分类器和 Snort 的网络入侵检测系统[32],目的在于在云环境中以低的误报率和负担得起的计算成本进行网络入侵检测。Shivaji 研究如何在无线传感器网络中以较低的能源消耗进行入侵检测,从而延长网络的生命周期[33]。其中贝叶斯方法用于传感器节点的能量预测。将贝叶斯算法应用于入侵检测技术中可以很好地解决在提升入侵检测率和执行可行性方面的困难和障碍,但是由于贝叶斯算法本身存在"条件独立性假设"的前提条件,使得在多数分类环境下无法发挥最佳的分类性能。很多学者对贝叶斯算法进行合理

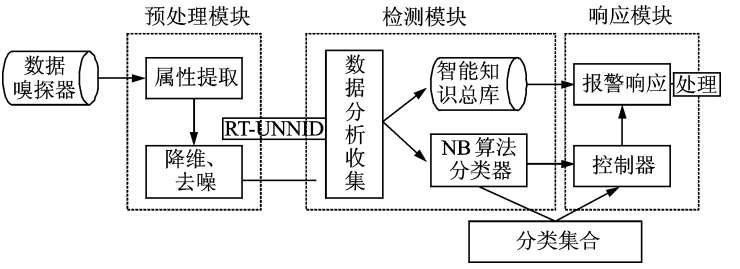

图 5　基于朴素贝叶斯(NB)的网络入侵模型

Fig. 5　Network intrusion model based on Naive Bayes(NB)

的改进后再应用到入侵检测模型之中。Klassen 提出了基于概率的贝叶斯分类器[34]。Xiao Liyuan 提出了一种贝叶斯网络模型平均(Bayesian network modeling average, BNMA)分类器[35],是基于 K-best 贝叶斯分类器用贝叶斯模型平均法建立贝叶斯分类器。Fariba Younes Nia 提出了一种应用于网络入侵检测并且可以有效提高检测质量的模型算法[36],在该算法中,结合了树增广朴素贝叶斯和 C5.0 决策树。Xiaoyan Han 等提出了一种基于主成分分析的朴素贝叶斯网络入侵检测算法[37],有效降低了数据维数,提高了检测的效率。

1.5 K-means 聚类

K-means 聚类[38]用于将数据识别到不同的类(被称为簇),通过不断获取离种子点最近的均值来计算数据聚集,是典型的基于距离的聚类算法。两个点的距离越近,其相似性越高,算法的最终目的是获得独立紧凑的簇。K-means 聚类的算法流程是[39]:

(1)随机选择 k 个对象作为初始聚类中心(种子点);

(2)计算每个点到中心对象的距离,并根据最小距离重新对相应点进行划分;

(3)移动聚类中心到属于它的"点群"中心;

(4)重复步骤(2,3),直到每个聚类不再发生变化。

聚类属于无监督学习,与上述的贝叶斯、支持向量机等含有类别标签的算法不同,聚类的样本中并没有给出标签。聚类是将对象分组成有意义的子类对象的方法,同一个集合的成员性质相似,不同集合的成员性质不同,因而聚类方法可以用于日志数据和入侵检测的分类[40]。Wang Huai bin 等将 SOM 网络与 K-means 相结合,利用 SOM 获得大致集群和集群中心,然后用 K-means 改进集群[41]。Muda Z 等将 K-means 和专家分类技术结合,提高了检测系统的正确性和检测率[42]。Liu Zhong 等提出了一种基于粒子群优化算法的 K-means 算法入侵检测方法[43],在仿真实验中取得了较好的效果。为了实现无监督入侵检测,Jirachan 等提出了在大型数据集识别异常值的可伸缩的孤立点检测方法[44],该方法结合了 Kolmogorov-Smirnov 检验和 K-means 聚类方法。K-means 算法简单、易于操作、复杂度低且处理数据十分快捷。但是由于算法本身的一些缺陷,会产生一些问题:初始聚类中心选取的不确定性、局部最小值、不适合形状延展性强的簇、K 值需要用户调节且对异常数据很敏感。针对算法中的缺陷,研究学者作出了部分改进,并用于网络入侵检测中。针对 K-means 算法的依赖和复杂性,Li Tian 通过研究传统聚类算法提出了一种改进的聚类算法[45],新算法学习了 k-Medoids 和改善三角形三边关系定理的优点。Pathak 提出了多线程 K-means 聚类方法[46]。Eslamnezhad 提出了基于极大极小 K-means 聚类的入侵检测方法[47],克服了传统 K-means 算法对初始中心敏感的缺陷,提高了聚类质量。Sandhya 在无线传感器网络中采用基于遗传算法的 K-means 算法进行异常检测[48],基于遗传算法的聚类提供了重要的减少输入特性的识别。K-means 算法在网络入侵检测系统的应用中,常常结合其他的机器学习算法提高检测系统的性能。Ashok[49]和 Pingjie Tang[50]两位作者在各自的论文中结合了支持向量机。Sharma[51]、Varuna[52]和 Muda[53]3 位作者在各自的论文中结合了贝叶斯。Chandrasekhar[54]等在原有的 K-means 算法的基础上结合了神经网络和支持向量机。Biswas[55]提出了一种新的入侵检测系统,该系统基于神经网络、K-means 聚类和主成分分析。聚类算法属于无监督学习,可以更好地解决无标签分类问题,算法本身简单易操作,处理数据快、速效率高,在网络入侵检测中,可以用于日志数据的检测和入侵检测的分类。

2 基于机器学习的网络入侵检测技术比较

面对日益复杂的网络环境,传统的网络入侵检测技术日渐乏力,急需新的技术来提高入侵检测系统的防御性能。将机器学习的算法应用于入侵检测系统中,提高了系统的检测效率,使系统更加智能化,

优化了整个系统的性能。机器学习中的算法多种多样,并各有利弊,表 1 简单介绍了各个算法的优缺点。基于决策树的网络入侵检测系统简单易理解,由于属性比较时只顺着一条支线,且构建的是最优决策树,所以大大提高了入侵检测系统的效率。基于神经网络的网络入侵检测系统主要优势在于高速并行计算。传统的入侵检测系统中央节点过大,并且缺乏主动防御能力而神经网络自学习联想记忆和高速并行计算的特性很好地弥补了这些问题。决策树和神经网络算法在一定程度上提高了入侵检测系统的性能,然而两种算法都容易陷入局部最小,出现过拟合的情况。基于支持向量机的入侵检测系统获得的是全局最小,具有较好的泛化能力。基于贝叶斯的入侵检测模型的主要优势在于概率推理能力,其对一个广泛的认知进行建模,这是其他算法所不具备的。将各个机器学习算法应用于网络入侵检测系统中时,无法断言某个具体的算法一定最佳,需要根据算法本身的优缺点和网络环境的特性进行选择,要具有针对性和适应性。

表 1　机器学习算法优缺点比较

Tab. 1　Advantages and disadvantages of machine learning algorithm

编号	机器学习方法	优点	缺点
1	决策树	大数据处理、高检测精度、模型清晰易理解。	建立决策树需要密集型计算,过度拟合。
2	神经网络	能在有限和不完整的数据中进行泛化研究,不需要专家知识。	训练过程较慢,不适用于实时检测;神经网络训练时可能产生过拟合。
3	支持向量机	样本数据小,较高的训练速度,可解决高维、非线性问题。	对缺失数据敏感,非线性问题无通用解决方法。
4	贝叶斯	对缺失数据不敏感,稳定的分类效率。	处理连续特征困难,需要知道先验概率。
5	K-means	计算快,输出易解释,聚类效果较好。	对异常值敏感,K 值难确定。

3　结束语

本文对机器学习算法在网络入侵检测中的应用进行了综述,简要介绍了机器学习中的常用算法,列举了不同算法在入侵检测中的应用。然而在网络入侵检测中最有效的方法并没有找到,并且考虑到方法的多样性和复杂性,不可能根据系统要检测的攻击类型只推荐一种方法。在决定方法有效性时,需要考虑多方面的因素,包括准确性、复杂性、用受过训练的模型对一个未知实例进行分类的时间和最终解决方案的可理解性。在一些特定的入侵检测系统中,需要针对特殊环境仔细考虑,建立一个健壮的网络入侵检测系统有待进一步研究。

参考文献:

[1] Butun I, Morgera S D, Sankar R. A survey of intrusion detection systems in wireless sensor networks[J]. IEEE Communications Surveys & Tutorials, 2014, 16(1):266-282.

[2] Huang M Y, Jasper R J, Wicks T M. A large scale distributed intrusion detection framework based on attack strategy analysis[J]. Computer Networks, 1998, 31(23/24):2465-2475.

[3] Mukkamala S, Sung A, Abraham A. Cyber security challenges: Designing efficient intrusion detection systems and antivirus tools[J]. Vemuri V Rao Enhancing Computer Security with Smart Technology, 2006, 5(12/2):125-163.

[4] Ayodele T O. Introduction to machine learning[C]// New Advances in Machine Learning. Rijeka, Croatia: [s. n.], 2010:1-20.

[5] Li J, Manikopoulos C N, Jorgenson J, et al. HIDE: A hierarchical network intrusion detection system using statistical preprocessing and neural network classification[C]// Proc IEEE Workshop on Information Assurance & Security. [S. l.]:IEEE, 2001:85-90.

[6] Hu W, Liao Y, Vemuri V R. Robust anomaly detection using support vector machines[C]// International Conference on Machine Learning. [S. l.]:Morgan Kaufmann Publishers, 2003:282-289.

[7] Landgrebe D. A survey of decision tree classifier methodology[J]. Systems Man & Cybernetics IEEE Transactions on,

1991, 21(3):660-674.

[8] Ayodele T O. Types of machine learning algorithms[J]. New Advances in Machine Learning, 2010,12(3):132-140.

[9] Komviriyavut T, Sangkatsanee P, Wattanapongsakorn N, et al. Network intrusion detection and classification with decision tree and rule based approaches[C]// International Symposium on Communications and Information Technology. Icheon:[s. n.],2009:1046-1050.

[10] Kbbes T, Bouhoula A, Rusinowitch M. Efficient decision tree for protocol analysis in intrusion detection[J]. International Journal of Security & Networks, 2010, 5(4):220-235.

[11] Sinapiromsaran K, Techaval N. Network intrusion detection using multi-attributed frame decision tree[C]// International Conference on Digital Information & Communication Technology & its Applications. Bangkok:[s. n.], 2012:203-207.

[12] Sahu S, Mehtre B M. Network intrusion detection system using J48 decision Tree[C]// International Conference on Advances in Computing, Communications and Informatics. Kochi:[s. n.] ,2015:1-6.

[13] Hornik K, stinchcombe M, White H. Multilayer feedforward networks are universal approximators[J]. Neural Networks, 1989, 2(5):359-366.

[14] Amini m, Jalili R, shahriari H R. RT-uNNid: A practical solution to real-time network-based intrusion detection using unsupervised neural networks[J]. Computers & Security, 2006, 25(6):459-468.

[15] Carpenter G, Grossberg S. Adaptive resonance theory[R]. CAS/CNS Technical Report Series-008, Cambridge, MA: MIT, 2003:89-90.

[16] Aljurayban N S, Emam A. Framework for cloud intrusion detection system service[C]//IEEE World Symposium on Web Applications and Networking. Sousse:IEEE, 2015:174-184.

[17] Ishitaki T, Elmazi D, liu Y, et al. Application of neural networks for intrusion detection in Tor networks[C]// IEEE, International Conference on Advanced Information Networking and Applications Workshops. Gwangiu:IEEE, 2015:67-72.

[18] Buczak A L, Guven E. A survey of data mining and machine learning methods for cyber security intrusion detection[J]. IEEE Communications Surveys & Tutorials, 2015, 99:1-26.

[19] Bao Xiaohui, Xu Tianqi, Hou Hui. Network intrusion detection based on support vector machine[C]// International Conference on Management and Service Science. Wuhan:[s. n.],2009:21-23.

[20] Li Huike, Gu Daquan. A novel intrusion detection scheme using support vector machine fuzzy network for mobile ad Hoc networks[C]// Web Mining and Web-based Application. Wuhan:[s. n.],2009:47-50.

[21] Li Lei, Gao Zhiping, Ding Wenyan. Fuzzy multi-class support vector machine based on binary tree in network intrusion detection[C]// International Conference on Electrical and Control Engineering. Wuhan:[s. n.],2010:1024-1046.

[22] Zhang Wei, Teng Shaohua, Zhu Haibin, et al. Fuzzy multi-class support vector machines for cooperative network intrusion detection[C]// IEEE International Conference on Cognitive Informatics. Beijing:IEEE,2010:811-818.

[23] Guo Jiangwei, Wu Xiaoping, Ye Qing. Network fault diagnosis based on rough set-support vector machine[C]// International Conference on Computer Application and System Modeling. Taiyuan:[s. n.],2010:312-315.

[24] Liu Zhiguo, Kang Jincui, Li Yuan. A hybrid method of rough set and support vector machine in network intrusion detection [C]// Signal Processing Systems (ICPS). Dalian:[s. n.],2010:561-563.

[25] Lin Nan, Xiang Chunzhi. A wavelet transform based support vector machine ensemble algorithm and its application in network intrusion detection[C]// Fifth International Conference on Intelligent Systems Design and Engineering Applications. Hunan:[s. n.], 2014:109-113.

[26] Lin Lizhong, Zhang Yaming, Zhang Yubin. Network intrusion detection method by least squares support vector machine classifier[C]// IEEE International Conference on Computer science and Information Technology. Chengdu:IEEE, 2010:295-297.

[27] Wang Zhi. Fault diagnosis for wireless sensor network based on genetic-support vector machine[C]// International Conference on Computer Science and Network Technology. Harbin:[s. n.], 2011:2691-2694.

[28] Hu Jianhong. Network intrusion detection algorithm based on improved support vector machine[C]// International Conference on Intelligent Transportation, Big Data and Smart City. Halong Bay:[s. n.], 2015:523-526.

[29] Luo Yangxia. The research of Bayesian classifier algorithms in intrusion detection system[C]// The International Conference on E-Business and E-Government, ICEE. Guangzhou, China:[s. n.], 2010:2174-2178.

[30] Fung C J, Zhang Jie, Boutaba R. Effective acquaintance management based on Bayesian learning for distributed intrusion detection networks[J]. IEEE Transactions on Network & Service Management, 2012, 9(3):320-332.

[31] Guan Keqing, Kong Xianli. Research on application of Bayesian discriminant method in intrusion detection model[C]// International Conference on Information Technology and Electronic Commerce. Dalian:[s. n.], 2014:180-183.

[32] Modi C N, Patel d R, Patel A, et al. Bayesian classifier and snort based network intrusion detection system in cloud computing[C]// Computing Communication & Networking Technologies (ICCCNT). Coimbatore:[s. n.], 2012:1-7.
[33] Shivaji S S, Patil A B. Energy efficient intrusion detection scheme based on Bayesian energy prediction in WSN[C]// 2015 Fifth International Conference on Advances in Computing and Communications (ICACC). Kochi:[s. n.], 2015:114-117.
[34] Klassen M, Yang N. Anomaly based intrusion detection in wireless networks using Bayesian classifier[C]// IEEE Fifth International Conference on Advanced Computational Intelligence. [S. l.]:IEEE,2012:257-264.
[35] Xiao Liyuan, Chen Yetian, Chang C K. Bayesian model averaging of Bayesian network classifiers for intrusion detection [C]// IEEE 38th International Computer Software and Applications Conference Workshops. Vasteras:IEEE, 2014:128-133.
[36] Nia F Y, Khalili M. An efficient modeling algorithm for intrusion detection systems using C5.0 and Bayesian network structures[C]// International Conference on Knowledge-Based Engineering and Innovation. Tehran:[s. n.],2015:1117-1123.
[37] Han Xiaoyan, Xu Liancheng, Ren Min, et al. A naive Bayesian network intrusion detection algorithm based on principal component analysis[C]// International Conference on Information Technology in Medicine and Education. Huangshan:[s. n.],2015:325-328.
[38] Kanungo T, Mount D M, Netanyahu N S, et al. An efficient K-means clustering algorithm: Analysis and implementation [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2002, 24(7):881-892.
[39] Wang Shenghui. Research of intrusion detection based on an improved K-means algorithm[C]// Second International Conference on Innovations in Bio-inspired Computing and Applications. Shenzhen:[s. n.],2011:274-276.
[40] Meng Jianliang, Shang Haikun, Ling Bian. The application on intrusion detection based on k-means cluster algorithm[C]// International Forum on Information Technology and Applications. Chengdu:[s. n.], 2009:150-152.
[41] Wang Huaibin, Yang Hongliang, Xu Zhijian, et al. A clustering algorithm use SOM and K-means in intrusion detection [C]// The International Conference on E-Business and E-Government. Guangzhou, China:[s. n.], 2010:1281-1284.
[42] Muda Z, Yassin W, Sulaiman M N, et al. Intrusion detection with K-means clustering and OneR classification[J]. Journal of Information Assurance & Security, 2012,7(6):55-66.
[43] Xiao Lizhong, shao Zhiqing, Liu Gang. K-means algorithm based on particle swarm optimization algorithm for anomaly intrusion detection[C]// Intelligent Control and Automation. Dalian:[s. n.], 2006:5854-5858.
[44] Jirachan T, Piromsopa K. Applying KSE-test and K-means clustering towards scalable unsupervised intrusion detection [C]// International Joint Conference on Computer Science and Software Engineering. Songkhla:[s. n.],2015:43-49.
[45] Li Tian, Wang Jianwen, Research on network intrusion detection system based on improved K-means clustering algorithm [C]// International Forum on Computer Science Technology and Applications. Chongqing:[s. n.],2010:76-79.
[46] Pathak V, Ananthanarayana V S. A novel multi-threaded K-means clustering approach for intrusion detection[C]// IEEE International Conference on Computer Science and Automation Engineering. [S. l.]:IEEE,2012:34-38.
[47] Eslamnezhad M, Varjani A Y. Intrusion detection based on minmax K-means clustering[C]// International Symposium on Telecommunications. Tehran:[s. n.], 2014:804-808.
[48] Sandhya G, Julian A. Iintrusion detection in wireless sensor network using genetic K-means algorithm[C]// International Conference on Advanced Communication, Control and Computing Technologies. Ramanathapuram:[s. n.], 2014:1791-1794.
[49] Ashok R, Lakshmi A J, Rani G D V, et al. Optimized feature selection with K-means clustered triangle svm for intrusion detection[C]// International Conference on Advanced Computing. Chennai:[s. n.], 2011:23-27.
[50] Tang Pingjie, Jiang Rongan, Zhao Mingwei. Feature selection and design of intrusion detection system based on K-means and triangle area support vector machine[C]// International Conference on Future Networks. Sanya, Hainan:[s. n.],2010:144-148.
[51] Sharma S K, pandey P, Tiwari S K, et al. An improved network intrusion detection technique based on K-means clustering via naïve Bayes classification[C]// International Conference on Advances in Engineering, Science and Management, Nagapattinam. Tamil Nadu, India:[s. n.],2012:417-422.
[52] Varuna S, Natesan P. An integration of K-means clustering and naïve Bayes classifier for intrusion detection[C]// International Conference on Signal Processing, Communication and Networking. Chennai:[s. n.],2015.
[53] Muda Z, Yassin W, Sulaiman M N, et al. Intrusion detection based on K-means clustering and naïve Bayes classification [C]// International Conference on Information Technology in Asia. [S. l.]:IEEE, 2011:1-6.
[54] Chandrasekhar A M, Raghuveer K. Intrusion detection technique by using K-means, fuzzy neural network and SVM classifiers[C]// International Conference on Computer Communication and Informatics. Coimbatore:[s. n.], 2013:1-7.

[55] Shah F M, Biswas N A, Tammi W M, et al. Fp-aNk: An improvised intrusion detection system with hybridization of neural network and K-means clustering over feature selection by PCA[C]//International Conference on Computer and Information Technology. Dhaka:[s. n.],2015:317-322.

作者简介:

朱琨(1984-),男,教授,研究方向:5G,自组织网络,E-mail: zhukun@nuaa.edu.cn。

张琪(1991-),女,硕士研究生,研究方向:机器学习、网络优化。

第五届中国计算机学会大数据学术会议(CCF BigData 2017)征文通知

第5届中国计算机学会大数据学术会议(CCF BigData 2017)将于2017年10月13日至15日在广东深圳市举行。本届会议由中国计算机学会(CCF)主办,CCF大数据专家委员会与深圳大学联合承办。本届会议拟突出多学科的交叉融合。会议将邀请多位院士和国际大数据领域的顶级专家学者作大会特邀报告,还将组织专题论坛、青年论坛和分会场口头报告等多种形式的学术交流。会议特别设立最佳学术论文奖、最佳应用论文奖和最佳学生论文奖。

重要日期:

投稿截止日期:2017年6月30日　　录用通知日期:2017年8月31日

正稿提交日期:2017年9月15日　　会议举办日期:2017年10月13～15日

征稿范围包括但不限于以下方面:数据科学基础理论与方法、数据科学与大数据趋势与未来、大数据系统构架与基础设施、大数据采集与预处理技术、大数据存储管理模型、技术与系统、大数据并行计算模型、框架与系统、主流开源大数据系统优化与应用实践、大数据分析挖掘与智能计算方法与系统、高性能大数据学习架构、算法与系统、大数据可视化分析与计算、大数据共享开放技术方法与标准、大数据隐私与安全保护、大数据系统解决方案与工具平台、大数据行业与政府应用。

征文分中文与英文论文,作者自选。本次会议不设会议文集,所有会议录用论文都将推荐到期刊,由期刊进行独立的审稿以决定是否录用到期刊,审稿的时间进度由各期刊掌握,大部分录用通知可能会在会议之后发出。录用的中文论文按论文质量推荐到《计算机学报》《计算机研究与发展》《电子学报》《中文信息学报》《模式识别与人工智能》《数据采集与处理》《清华大学学报(自然科学版)》《南京大学学报(自然科学版)》《中国科学技术大学学报》《计算机科学与探索》《计算机科学》《计算机应用》《计算机工程与科学》《智能系统学报》《大数据》等国内核心期刊的正刊上发表。

录用的英文论文将根据论文质量推荐到 *IEEE Transactions on Big Data*, *International Journal of Machine Learning and Cybernetics* (SCI), *International Journal of Data Science and Analytics*, *Interdisciplinary Sciences*: *Computational Life Science* (SCI), *International Journal of Computational Science and Engineering* (EI), *International Journal of Embedded Systems* (EI), *Big Data Mining and Analytics*(EI)和 *International Journal of High Performance Computing and Networking* (EI)等英文期刊。

投稿咨询和支持: 庞观松　邮箱:ccfbigdata2017@gmail.com

会务咨询:崔来中　邮箱:ccfbigdata2017@szu.edu.cn　　电话:0755-26906581-804