贝叶斯分类在入侵检测中的应用研究

薛静锋　曹元大
（北京理工大学软件学院　北京 100081）

摘　要　根据分类技术建立入侵检测模型的思路，构造了一个基于贝叶斯分类的入侵检测原型系统。为了解决该方法存在的训练数据集问题，本文改进了现有的贝叶斯分类算法，提出了利用未标记数据提高贝叶斯分类器性能的方法。实验表明，这种方法取得了很好的效果。

关键词　网络安全，入侵检测，贝叶斯分类

The Application of Bayes Classification in IDS

XUE Jing-Feng　CAO Yuan-Da
(School of Software, Beijing Institute of Technology, Beijing 100081)

Abstract　According to the idea of designing intrusion detection system using classification, an intrusion detection system prototype is developed. In order to solve the problem existing in training data sets, present Bayes algorithm is improved and an algorithm using unlabeled data to improve the capability of the classifier is proposed. The experiment shows the result of this algorithm is good.

Keywords　Network security, Intrusion detection, Bayes classification

1　入侵检测的分类思想

1999 年，Wenke Lee 在其博士论文中给出了用数据挖掘技术建立入侵检测模型的过程[1]，如图 1 所示。

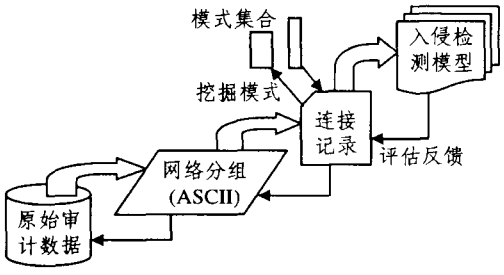

图 1　用数据挖掘建立入侵检测模型的过程

在图 1 中，原始审计数据是从网络或者主机上获取的二进制的审计数据。首先把原始审计数据转换成 ASCII 格式的网络分组信息，再把网络分组信息经过数据预处理程序处理成连接记录。其次，用数据挖掘算法中的关联分析算法和序列分析算法挖掘连接记录数据库中的频繁模式，如关联规则和频繁序列。利用这些频繁模式，为连接记录构造附加特征，如时间统计特征。最后，进行入侵检测模型的构建，数据挖掘有很多模型和算法，其中大部分算法都不是专为解决某个问题而特制的，算法之间也不相互排斥。在这些算法中，有几种算法对于分析网络审计数据和检测入侵非常有用，它们是关联规则算法、频繁序列算法和分类算法。利用这些算法及其它工具，可以从收集到的原始审计数据中建立入侵检测模型。

在上述三个步骤中，最核心的部分是第三步，即入侵检测模型的建立。对于入侵检测模型的建立，由于分类算法理论比较成熟，而且在实际应用中效果也比较好，因此得到了比较广泛的应用。

入侵检测的分类思想的基本思路是[1,2]：使用带标记的连接记录数据（训练数据）对分类器进行训练学习，该训练过程可能需要不断地反复和评估，如果分类器的分类效果不好，就需要通过频繁模式的挖掘和比较，构造更有助于分类的特征项。训练完成后的分类器就可以用于检测过程，将当前连接记录输入给分类器，分类器将输出该连接记录所属的类别。

2　基于贝叶斯分类的入侵检测系统的建立

利用 Wenke Lee 提出的用数据挖掘建立入侵检测模型的过程和分类模型的基本思路，本文构造了一个基于贝叶斯分类的入侵检测原型系统。由于目前基于数据挖掘入侵检测的最大弊端是实时性不高，因此本文在选择分类算法时选择了贝叶斯分类算法，该算法具有实现简单、分类速度快，利于实时检测的优势。

2.1　系统结构

本文构造的基于贝叶斯分类的入侵检测原型系统的基本结构如图 2 所示。整个系统的工作过程由训练过程和检测过程组成。在训练过程中，系统使用大量网络连接记录组成的带标记的训练数据集，对贝叶斯分类器进行训练，通过不断循环反馈使得分类器可以分辨或预测哪些行为是正常的，哪些行为是不正常的。在检测过程中，系统利用训练过程中得到的知识库，使用训练好的贝叶斯分类器对当前连接记录进行分类，从而判断出当前行为是正常行为还是异常行为。

2.2　训练过程

在训练过程中，数据采集模块采用的是 tcpdump 程序[3]。tcpdump 截获的数据不能直接用分类算法进行分析，因此首先需要对它进行预处理，从中提取有意义的特征。数据预处理模块负责对数据采集模块采集到的 tcpdump 格式的数据进行预处理，以便把它们转换为连接记录的形式。对

薛静锋　讲师，博士，主要研究方向为网络安全技术。**曹元大**　教授，博士生导师，主要研究方向为网络安全技术和人工智能技术。

每一条 TCP 连接，把所有关于一条连接的所有数据包整理成一条连接记录。因为 UDP 是面向无连接的数据传输协议，在数据传输过程中不存在握手连接或拆除连接的过程，所以把每个 UDP 数据包看成是一条连接[2]。此外，把每个 ICMP 包也看作是一条连接。

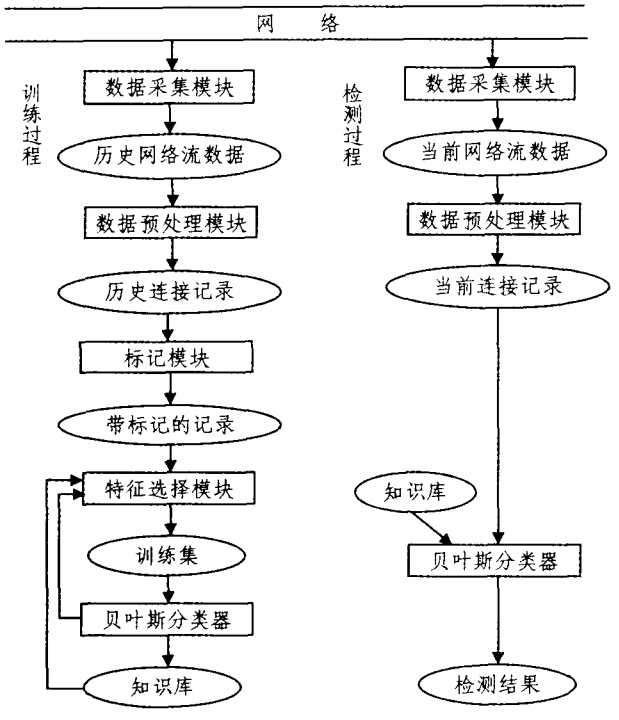

图 2　基于贝叶斯分类入侵检测原型系统

标记模块的作用是对训练数据进行标记，以便区分出正常记录和攻击记录。一般来说，标记都是通过手工来完成的。

将带标记的连接记录输入给特征选择模块。由于连接记录可能包含很多特征，而在检测中并不需要分析全部特征，即会有冗余的特征存在，因此应该把这些冗余特征去掉。

使用训练集对贝叶斯分类器进行训练，在训练过程中，需要依据训练结果指导特征选择模块进行更进一步的特征选择，不断优化所采用的特征集合，该过程循环往复，直到得到良好的分类结果为止。从而形成稳定的知识库。该知识库将用于检测过程。

2.3　检测过程

在检测过程中，数据采集模块从网络上采集当前网络流数据，并通过数据预处理模块将其转换为当前连接记录。这两个模块的工作原理与前述相同，在此不再赘述。

当前连接记录被送到贝叶斯分类器中，由贝叶斯分类器依据相应的知识库对其进行分类检测，从而决定是否有违反安全策略的入侵行为发生。

2.4　实验设计

为了验证上述入侵检测原型系统的有效性，我们设计了如下实验并进行了实验分析。

实验使用的训练数据和测试数据摘自 http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html，这是第五届知识发现和数据挖掘国际会议(KDD-99)为测试基于网络的入侵检测系统所提供的数据，该数据具有以下特点：①每条记录由 41 个特征(属性)构成，其中有 34 个属性为连续值，7 个属性为离散值。②每条记录都被标记为正常(normal)或是一种入侵行为(如 smurf)。

为了实验的方便，我们抽取了部分数据形成数据集 D。D 中含有 9 种攻击类型的数据：land、neptune、pod、teardrop、nmap、portsweep、satan、mscan 和 ipsweep，通过对这 9 种攻击方式进行分析，我们确定了最终使用的特征集合。

在实验中，分别选取两个不同的训练数据集 $T1$ 和 $T2$ 分别对贝叶斯分类器 $C1$ 和 $C2$ 进行训练，然后分别用训练好的 $C1$ 和 $C2$ 对测试数据集 E 进行分类。其中 $T1$、$T2$、E 和 D 之间满足以下关系：

$T1 \subset D, |T1| = 10,000$

$T2 \subset D, |T2| = 100,000$

$T1 \subset T2$

$E \subset D, |E| = 100,000$

并且，$T1$ 和 $T2$ 中只包含有 D 中除 mscan 和 ipsweep 之外的 7 种攻击方式，而 E 中包含有 D 中全部的 9 种攻击方式。

实验的输出为贝叶斯分类器 $C1$ 和 $C2$ 的检测率，而误报率则通过分类器对正常记录(normal)的检测率来体现。具体实验结果如表 1 所示。

表 1　检测率

检测率(%)	normal	land	neptune	pod	teardrop	nmap	portsweep	satan	mscan	ipsweep
C1	95.95	99.33	91.96	88.11	97.00	87.05	94.94	81.78	72.33	62.56
C2	98.81	100	96.90	95.40	99.50	92.81	99.72	90.96	82.63	79.41

2.5　实验分析

为了直观起见和便于分析，把表 1 所示的实验结果用柱形图表示出来，如图 3 所示。

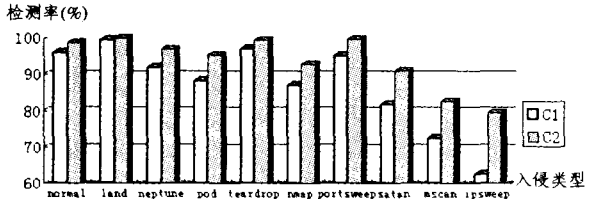

图 3　实验结果的柱形图表示

根据表 1 和图 3 所示的实验结果，可以得到如下结论：

①贝叶斯分类对于基于网络的入侵检测是有效的：对于 $T1$ 和 $T2$ 中的 7 种入侵攻击类型，$C1$ 和 $C2$ 的检测率均在 80% 以上，个别时候能达到 100%。

②贝叶斯分类对于异常入侵检测也是有效的：对于 $T1$ 和 $T2$ 所不包含的 2 种入侵攻击方式 mscan 和 ipsweep，$C1$ 和 $C2$ 的检测率能达到 60%以上，最高能达到 80%多。

③分类器的检测效率与训练数据集的大小密切相关：从图 3 可以看出，$C2$ 的检测率明显好于 $C1$，原因在于 $C2$ 的训练数据集明显大于 $C1$ 的训练数据集。

2.6　存在的问题

通过分析结果可以看出，贝叶斯分类对于基于网络的入侵检测是有效的；但是通过对上述结果③的深入分析，也发现了基于贝叶斯分类的入侵检测技术存在的重要问题：训练数据集问题。分析结果③表明，$C2$ 的检测率明显好于 $C1$，原因在于 $C2$ 的训练数据集明显大于 $C1$ 的训练数据集。因此通过扩大训练集可以提高分类器的检测率，但是该方法并没有我们想象的那样容易。训练集中的连接记录必须事先经过标记，而这种标记必须由人手工完成，这是一项费时费力的工作。人们更喜欢仅需要手工标记少量连接记录而不是大量连接记录，例如，$C2$ 的检测效率虽然高于 $C1$，但 $T2$ 中的连接记录数量却是 $T1$ 的 10 倍，达到了 100,000 条，在具体实现中这是很不现实的。因此，应该寻求一种解决方法，使得通过训练少量带标记的连接记录数据就可以获得较好的检测率。

3 利用未标记数据提高贝叶斯分类器性能的方法

为了解决上述问题，本文改进了现有的贝叶斯分类算法，提出了利用未标记数据提高贝叶斯分类器性能的方法。

3.1 基本思想

本文提出的利用未标记数据提高贝叶斯分类器性能的基本思想是：设有两类数据 CN（正常类）和 CA（入侵类），首先用少量的带标记的训练数据（数据集 $D1$）对贝叶斯分类器进行训练，得到分类器 C，然后依次将大量的未带标记的数据（数据集 $D2$）输入给 C，由 C 对每条未带标记的连接记录 r 进行分类，因为贝叶斯分类的原理是计算 r 分别属于 CN 和 CA 的概率[4]，设 r 属于 CN 的概率为 PN，属于 CA 的概率为 PA，如果 $PN>PA$，则认为 r 属于 CN，否则认为 r 属于 CA。为此，我们的思路是，设定一个阈值 $e(e>1)$，对于连接记录 r，如果 $PN/PA>e$，即 r 属于类 CN 的概率远大于 r 属于类 CA 的概率，则将 r 归为类 CN；如果 $PA/PN>e$，即 r 属于类 CA 的概率远大于 r 属于类 CN 的概率，则将 r 归为类 CA；如果 PN 和 PA 不符合上面两个条件，即不能判断 r 明确属于哪个类，这时通过人工干预进行判断，确定 r 的明确归属。通过这种方式给大量未带标记的数据进行自动标记，然后继续对贝叶斯分类器进行训练，以得到检测性能更好的分类器。

3.2 利用未标记数据提高贝叶斯分类器性能的算法

根据上述利用未标记数据提高贝叶斯分类器性能的基本思想，本文设计了利用未标记数据提高贝叶斯分类器性能的算法，如下所示：

算法：利用未标记数据提高贝叶斯分类器性能的算法

输入：未训练的贝叶斯分类器 C；判断概率差的阈值 e；带标记的训练数据集 $D1$；未带标记的训练数据集 $D2$；

输出：训练好的贝叶斯分类器 C；知识库 K。

方法：

(1)用带标记的数据集 $D1$ 对贝叶斯分类器进行训练，得到分类器 C 和知识库 K；
(2)for($D2$ 中的每一条连接记录 r)
(3)用 C 和 K 对 r 进行分类，计算 PN 和 PA；
(4)如果 $PN/PA>e$，则 r 为正常数据，TC 中相应计数器加 1；
(5)否则，如果 $PA/PN>e$，则 r 为入侵数据，TC 中相应计数器加 1；
(6)否则，r 需人工判断，将 r 保存于人工判断数据集 $D3$，转(2)；
(7)计算新的 C 和 K；
(8)end for；
(9)for($D3$ 中的每一条连接记录 r)
(10)人工判断 r 属于何种类型数据；
(11)计算新的 C 和 K；
(12)end for；
(13)return C 和 K。

该算法首先用带标记的数据集 $D1$ 对贝叶斯分类器进行训练，得到分类器 C，然后将分类器 C 作用于未带标记的训练数据集 $D2$，对 $D2$ 中的每一条连接记录 r，用 C 对其进行分类（也可能需要人工干预），然后根据分类结果训练分类器 C，得到新的 C 和 K。

3.3 实验分析

为了验证算法的有效性，我们在 1.4 节实验的基础上，设计了如下实验：

算法的输入为：$D1=T1$；$D2=T2-T1$；$e=2.8$。其中 $T1$ 与 $T2$ 如先前所描述，也就是说，这里的 $|D1|=10,000$，$|D2|=90,000$，并将 $D2$ 中每条记录的标记都撤掉。经过 $D1$ 与 $D2$ 对贝叶斯分类器的训练，算法输出训练好的分类器 C 和知识库 K。然后，用该分类器 C 对先前的测试数据集 E 进行检测，得到的检测结果如表 2 所示。

表 2 检测结果

检测率(%)	normal	land	neptune	pod	teardrop	nmap	portsweep	satan	mscan	ipsweep
C	97.15	100	93.33	94.27	99.50	90.76	96.88	87.67	79.13	69.76

为便于比较，把表 2 中的数据与表 1 中的数据一起以柱形图形式表示出来，如图 4 所示。

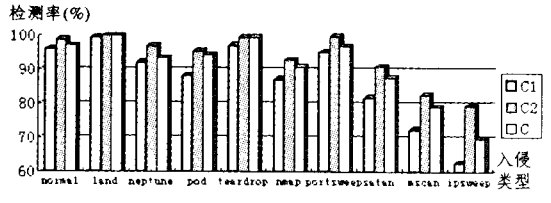

图 4 实验结果的柱形图表示

从图 4 中可以看出，C 的检测率虽然不及 $C2$，但要明显好于 $C1$，而且 C 对两种入侵类型（land 和 teardrop）的检测率与 $C2$ 相同。此外，在该实验中，通过人工干预进行判断的连接记录数为 919 条，人工干预率为 (919/90,000) * 100%=1.02%，我们认为，这对于用户来说是可以接受的，而且，通过调节参数 e，还可以进一步降低人工干预率。

以上实验分析表明，利用未标记数据提高贝叶斯分类器性能的方法是有效的，在使用少量带标记的训练数据对贝叶斯分类器进行训练的基础上，通过对大量未带标记的训练数据进行自动标记从而继续对其进行训练，可以提高贝叶斯分类器的分类性能，从而较好地解决了训练数据集问题。

结束语 本文根据 Wenke Lee 提出的用基于分类的数据挖掘技术建立入侵检测模型的思路，构造了一个基于贝叶斯分类的入侵检测原型系统。实验表明，该系统对于入侵检测是有效的，而且具有异常检测的能力。但实验结果也反映出该方法存在着训练数据集问题：通过扩大训练集可以提高分类器的检测率，但是训练集中的连接记录必须事先经过标记，而这种标记必须由人手工完成，这是一项费时费力的工

一种新的自适应盲数字水印算法*)

彭求明 杨小帆 黄 松 李思静 柏 森
(重庆大学计算机学院 重庆 400044) (重庆通信学院研究生管理大队 重庆 400035)

摘 要 本文提出了一种新颖的基于DCT域的自适应盲数字水印算法,该算法由三个阶段组成。首先,根据人眼视觉系统(HVS)和图像的局部特性来选择水印嵌入区域;然后充分利用JPEG压缩量化阶段的舍入误差;第三个阶段是采用一组公式来嵌入多个版本的水印,水印是经位扩展和二值随机序列调制得到的。水印的检测不需要原始图像。实验结果表明,该算法在保证水印不可见性的同时,对常见的几种图像处理如缩放、椒盐噪声、滤波特别是JPEG压缩和剪切有很高的鲁棒性。

关键词 数字水印,离散余弦变换(DCT),人眼视觉系统(HVS),JPEG压缩,盲检测

A Novel Adaptive Blind Digital Watermarking Algorithm

PENG Qiu-Ming YANG Xiao-Fan HUANG Song LI Si-Jing BAI Sen
(College of Computer, Chongqing University, Chongqing 400044)
(Graduate Education Division, Chongqing Communication Institute, Chongqing 400035)

Abstract In this paper, an adaptive blind digital watermarking scheme is proposed. The algorithm consists of three phases. First, the embedding region of watermark is selected via the human visual system model and local characteristics of image; then the round-off errors is analysed during the quantization step of JPEG compression; at the third phase, several simple formulas are assumed to embed several watermarks that are generated by bit-spreading and binary random sequence's modulation. The watermark is detected without any reference to the original image. Experimental results show that the embedded watermark is invisible; while at the same time the algorithm is robust to some typical kinds of image processing such as scaling, salt & pepper noise, filtering, especially, JPEG compression and cropping attacks.

Keywords Digital watermarking, DCT, HVS, JPEG compression, Blind detection

1 引言

近年来,随着互联网和多媒体信息处理技术的迅猛发展,网上多媒体信息(图像、视频、音频、文本等)的免费获取、复制、分发、传送、修改已变得极为快捷、方便。然而,网络给人们带来便利的同时也暴漏了越来越多的安全问题,如何防止非法复制、有效地保护版权已成为一个极为迫切的问题。数字水印技术是一种有效的保护版权的手段,已经引起了人们的高度重视和广泛关注,数字水印技术的实现主要考虑两个因素:稳健性和不可见性[1]。这两者是相互矛盾的,一个可靠的数字水印系统应该是稳健性和不可见性的最佳折衷。

数字水印技术一般可以分为两类:空域水印[2]和变换域水印。空域水印直接修改图像的象素,该类算法的最大特点是算法简单、计算复杂度低,但鲁棒性差。变换域算法是对图像进行各种变换(常见的变换有DCT[3~6]、DWT[7]、DFT)后嵌入水印。相对于空域水印算法,变换域水印算法有如下优点:(1)在变换域内嵌入的水印信号能量可以分布到空间域的所有像素上,有利于提高鲁棒性;(2)可以较方便地结合HVS,有利于保证水印的不可见性;(3)变换域算法与现今大多数国际图像和视频压缩标准兼容,可直接实现压缩域水印嵌入。正因如此,变换域水印是当今水印技术研究的主流。

许多文献[8~13]提出了自适应水印算法,它们都是利用HVS和图像自身的特点自适应地嵌入水印,自适应主要体现在:嵌入位置的选取、嵌入水印强度的调节、嵌入水印位置个数的选择。文[8]根据空域8×8图像子块边缘点密度把所有的图像子块分为两类:弱纹理和强纹理,水印嵌在每一DCT系数块的DC(direct component)分量上,利用一组公式来量化修改它,通过拉伸因子来调节水印嵌入的强度,强纹理块用大的拉伸因子,弱纹理块用小的拉伸因子,是盲检测。文[9]是根据亮度掩蔽特性和纹理掩蔽特性把所有的8×8DCT子块分为两类:适合嵌入水印和不适合嵌入水印。由DC分量的大小来评价亮度特性;把每一个DCT系数块用JPEG压缩量化矩阵来量化,根据非零个数的多少来判断纹理的复杂程度。水印嵌在中频区较大的系数上,不同的块嵌入不同能量

*)基金项目:本文工作受到重庆应用基础研究项目(编号:8028)资助。彭求明 硕士研究生,研究方向:数字水印、图像处理。杨小帆 教授,博导,研究方向:并行计算、容错及故障诊断、人工神经网络、数理方程。

作,对大量数据进行手工标记在具体实现中是很不现实的。为了解决该问题,本文改进了现有的贝叶斯分类算法,提出了利用未标记数据提高贝叶斯分类器性能的方法。实验表明,这种方法取得了很好的效果。

参 考 文 献

1 Lee W. A data mining framework for constructing features and models for intrusion detection systems:[dissertation of Doctor of Philosophy]. Columbia University, 1999
2 Lee W, Stolfo S J. Data mining approaches for intrusion detection. the Seventh USENIX Security Symposium (SECURITY '98), San Antonio, TX, Jan. 1998
3 Jacobson V, Leres C, McCanne S. tcpdump. Available via anonymous ftp to ftp. ee. lbl. gov, June 1989
4 薛静锋,曹元大. 基于贝叶斯分类的分组入侵检测技术研究. 见:第三届全国CSCW暨第一届全国AIN学术会议论文集, 2002