



云环境中改进 FCM 和规则参数优化的网络入侵检测方法

张春琴^{1,2}, 谢立春¹

(1. 浙江工业职业技术学院, 浙江 绍兴 312000; 2. 浙江工业大学, 浙江 杭州 310014)

摘要: 针对云环境中的网络入侵检测问题, 提出一种基于模糊推理的网络入侵检测方法。首先, 利用互信息特征选择对样本特征进行降维。然后, 利用提出的改进模糊 C 均值聚类 (IFCM) 方法对训练样本集进行聚类, 根据各样本特征与集群的对应关系获得初始模糊规则库。接着, 对每个规则的前件参数和后件参数进行调优, 以此获得准确的规则库。最后, 基于规则库对输入连接数据进行模糊推理, 对其进行分类以实现入侵检测。在云入侵检测数据集上的实验结果表明, 该方法能够准确检测出网络入侵, 具有可行性和有效性。

关键词: 云环境; 网络入侵检测; 互信息特征选择; 改进模糊 C 均值聚类; 模糊规则库优化

中图分类号: TP393

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2018005

Network intrusion detection method based on improved FCM and rule parameter optimization in cloud environment

ZANG Chunqin^{1,2}, XIE Lichun¹

1. Zhejiang Industry Polytechnic College, Shaoxing 312000, China

2. Zhejiang University of Technology, Hangzhou 310014, China

Abstract: Aiming at the network intrusion detection problem in cloud environment, a method of network intrusion detection based on fuzzy inference was proposed. Firstly, it used the mutual information feature selection to reduce the feature of the sample. Then, the improved fuzzy C-means clustering method was used to cluster the training sample set, and the initial fuzzy rule base was got by the correspondence between each sample feature and cluster. After that, the refine parameter and consequent parameters of each rule were tuned to obtain an exact rule base. Finally, fuzzy inference was carried out on the input connection data based on the rule base, and it was classified to realize intrusion detection. Experimental results on the cloud intrusion detection dataset show that this method can detect the network intrusion accurately, and it is feasible and effective.

Key words: cloud environment, network intrusion detection, mutual information feature selection, improved fuzzy C-means clustering, fuzzy rule base optimization

收稿日期: 2017-06-22; 修回日期: 2017-09-25

基金项目: 国家自然科学基金青年科学基金资助项目 (No.61603211)

Foundation Item: The Young Science Foundation of National Natural Science Foundation of China (No.61603211)

1 引言

在云计算环境中,用户通过远程连接来访问所需的计算、存储、应用等资源,且数据通常存放在共享环境中。在实际情况中可能会存在一些恶意用户对数据进行窃取或篡改。为了提高云环境中的数据安全性,有必要对云平台中的网络用户访问数据流进行实时主动的监控和防御^[1]。目前,网络入侵检测系统(intrusion detection system, IDS)主要可分为两类^[2]:基于数据挖掘的IDS和基于机器学习的IDS。其中,机器学习方法是基于对数据的智能学习来构建分类器,其检测性能较好,但计算量较大,很难满足实时检测的要求^[3]。基于数据挖掘的IDS是通过对历史数据进行数据分析,获得一定的规律来检测入侵。其结构简单、执行速度快,然而其基于一些固定的规则,适应能力较差。为此,学者在数据挖掘中融入了模糊推理^[4],即通过训练数据集挖掘出网络连接数据特征与类别相对应的模糊规则,以此对输入连接数据进行推理来判别是否为入侵行为。

对于基于模糊推理的入侵检测方法,其有效性很大程度上依赖于模糊规则库。模糊规则库的构建通常分为基于神经网络和基于聚类技术的方法^[5]。其中,基于神经网络的方法结构复杂,且与网络表达的规则存在冲突。基于聚类的方法是通过将样本数据集进行聚类,获得的每个集群对应一个模糊规则。 K 均值(K -means)聚类算法^[6]是常用的聚类算法,然而,其聚类中心数量和位置的初始值直接影响聚类效果。为此,参考文献[7]提出了一种改进 K -means聚类算法,用来获得模糊规则,其通过一个半径函数来自适应选择聚类中心,一定程度上解决了初始值的问题。参考文献[8]在初始聚类中融入了模糊理论,提出了一种基于模糊 C 均值(fuzzy C -means, FCM)聚类算法的规则构建方法,利用模糊 C 均值聚类对训练样本进行聚类,根据各集群来生成模糊规则。然

而,传统FCM仍然存在初始聚类中心选择的问题。另外,现有的基于聚类技术获得模糊规则的方法中,通常获得的模糊规则较为粗糙,为此需要对其进行细化,以提高规则的准确性。

基于上述分析,提出一种新型的模糊规则生成方法,并将其应用到云环境的入侵检测中。首先,根据训练样本聚类构建规则库。然后,基于获得的模糊规则库,对网络连接数据进行推理,以判定其类别,实现入侵检测。提出方法的主要创新点在于:提出一种改进型FCM(improved FCM, IFCM)聚类算法,融于一个聚类中心近似度量因子来解决传统FCM聚类中的初始聚类中心选择问题,以此对样本进行精确聚类,构建初始模糊规则集;为了提高规则的准确性,提出了一种规则优化方法,即通过调优每个规则的前件参数和后件参数来对其进行优化。在一个最新的云入侵检测数据集(cloud intrusion detection dataset, CIDD)中的实验表明,提出方法的检测率能够达到98%以上,具有有效性。

2 提出方法的框架

本文提出一种用于云环境中网络入侵的检测系统,可部署在实际云平台中各物理服务器上,用来监控用户远程访问。检测系统分为训练和测试阶段,其基本框架如图1所示。

训练阶段主要用来获得模糊规则。首先,由于原始数据集中的特征较多且存在冗余,所以先通过基于互信息的特征选择方法来降低特征维度。然后,利用提出的模糊聚类方法对训练数据进行聚类,用来确定模糊输入/输出的关系,从而获得模糊规则。接着,通过一个动态优化学习算法,根据规则所涉及的前件和后件参数来优化模糊规则,构建最终的模糊规则库。

在测试阶段中,根据收集的网络连接数据和模糊规则库,推理输入数据与各类的相似度来进行分类,以此实现入侵检测。

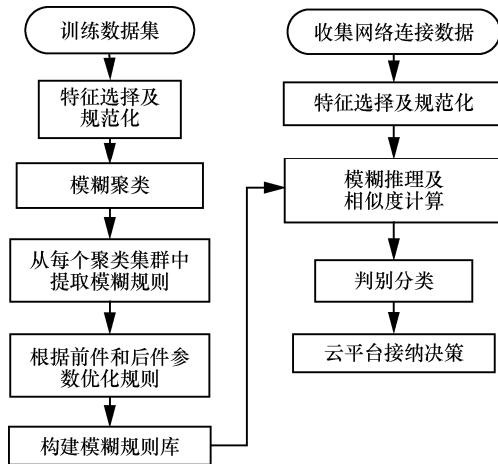


图1 提出入侵检测方法的基本框架

3 基于互信息的特征选择

网络连接数据具有大量的特征，冗余特征不仅增加了分类器的计算时间，而且还会降低检测准确率。为此，在输入分类器进行检测之前，需要对其进行特征降维。目前，特征选择方法主要有欧氏距离法、余弦相似度法、互信息（mutual information, MI）法等^[9]。其中，互信息法是衡量两个随机变量之间相关性的一种有效方法。

3.1 互信息

互信息是两个随机变量之间关系的对称度量，输出为一个非负值，其中，零表示两个变量是统计独立的^[10]。

给定 2 个联系随机变量 $U = \{u_1, u_2, \dots, u_d\}$ 和 $V = \{v_1, v_2, \dots, v_d\}$ ，其中， d 为样本总数， U 和 V 之间的 MI 定义为：

$$I(U;V) = H(U) + H(V) - H(U,V) \quad (1)$$

其中， $H(U)$ 和 $H(V)$ 为 U 和 V 的信息熵，信息熵为变量不确定性的度量，表达式为：

$$\begin{aligned} H(U) &= -\int_U p(u) \lg p(u) du \\ H(V) &= -\int_V p(v) \lg p(v) dv \end{aligned} \quad (2)$$

U 和 V 之间的联合熵定义为：

$$H(U;V) = -\int_U \int_V p(u,v) \lg p(u,v) du dv \quad (3)$$

那么， U 和 V 之间的 MI 可表示为：

$$I(U;V) = -\int_U \int_V p(u,v) \lg \frac{p(u,v)}{p(u)p(v)} du dv \quad (4)$$

对于离散化的 U 和 V ，则 MI 可表示为：

$$I(U;V) = \sum_{u \in U} \sum_{v \in V} p(u,v) \lg \frac{p(u,v)}{p(u)p(v)} \quad (5)$$

在特征选择中，如果特征包含关于类的重要信息，则该特征与该类相关；否则不相关或多余。由于 MI 可以量化两个随机变量共享的信息量，因此可用作评估特征和类标签之间相关性的标准。具有更高预测能力的特征 f 具有更大的互信息 $I(C;f)$ 。

3.2 互信息特征选择

基于互信息理论来选择特征的方法称为互信息特征选择（MIFS）^[11]。MIFS 中，其通过计算 $I(C;f_i,S)$ 来选择特征。 $I(C;f_i,S)$ 表示在所选择的特征集 S 中，增加特征 f_i 后形成的新特征集与类别 C 之间的互信息，反映了特征 f_i 对分类的贡献程度。 $I(C;f_i,S)$ 的计算思想是计算特征 f_i 与类别 C 之间的互信息 $I(C;f_i)$ ，然后计算特征 f_i 与先前所选特征 f_s 之间的互信息 $I(f_i;f_s)$ 之和，并将其乘以一个比例系数 β 对 $I(C;f_i)$ 进行校正。表达式如下：

$$I(C;f_i,S) = I(C;f_i) - \beta \sum_{f_s \in S} I(f_i;f_s) \quad (6)$$

β 值表示进行特征选择时考虑输入特征之间冗余度的权重，其决定着在选择特征时，两个方面（即特征与类之间的 MI 和特征与特征之间的 MI）的重要性比重。为了确定最佳的 β 值，设定 β 分别从 [0.1, 1] 范围内取值，步长为 0.1。在其他条件不变的情况下，在 CIDDD 数据集上采用下文提出的入侵检测方法进行了实验。结果表明，当 $\beta=0.3$ 时，检测性能最好，即所选择出的特征最优。为了进一步确定 β 值在不同数据集下的普适性，在传统的 KDD99 数据集上也进行了相关实

验,表明 β 取值为0.3和0.4时效果都较好,且入侵检测结果相差很小。为此,在本文中,统一设置 $\beta=0.3$ 。

4 提出的入侵检测方法

4.1 传统FCM聚类

FCM聚类算法是一种考虑样本模糊归属的软划分方法,其根据每个样本对所有聚类中心的隶属度进行自动聚类。

假设入侵检测数据集包含 K 个样本,形式为 $[I_1, O_1], [I_2, O_2], \dots, [I_K, O_K]$ 。其中, I_i 表示样本数据特征,每个 $I_i = [I_{i1}, I_{i2}, \dots, I_{iN}]$, N 为特征数。 O_i 为输出变量,表示5种类别之一,即4类攻击和正常访问。

FCM将 K 个样本分为 C 个集群,并求每个集群的中心 z_i ,目标函数定义为:

$$J(U, Z) = \sum_{i=1}^K \sum_{j=1}^C u_{ij} d_{ij}^2 \quad (7)$$

其中, u_{ij} 表示第 i 个样本对第 j 类的隶属度值; d_{ij} 表示第 i 个样本与第 j 个集群中心的距离,这里采用Minkowski(闵可夫斯基)距离度量。

FCM算法根据预设定的聚类数量 C ,通过迭代过程来最小化目标函数 J 。由于传统FCM随机选择初始聚类中心,且只考虑距离度量进行聚类,所以其聚类效果不是很理想。

4.2 改进型FCM聚类

针对传统FCM中存在聚类中心初始化的问题,提出一种用于聚类中心选取的方法。其中,定义了一个聚类中心近似度量 Q ,表示第 i 个样本 I_i 到其他任意一个样本的距离,表达式为:

$$Q_i = \sum_{j=1}^K e^{-\phi \|I_i - I_j\|^2} \quad (8)$$

其中, ϕ 为一个常数。那么,提出的改进FCM聚类算法的步骤如下。

步骤1 计算每个样本的 Q 值,选择具有最大 Q 值的样本 I_1^* 作为第一个聚类中心 c_1^* ,并将该

Q 值设定为 Q_1^* 。

步骤2 计算样本与该聚类中心的隶属度值,根据隶属度值,将 c_1^* 点和与 c_1^* 具有很高隶属度值的样本从下一次迭代中移除,防止产生伪聚类中心。

步骤3 通过再一次计算 Q 值来确定下一个聚类中心 c_1^* ,即 $Q_{i+1} = Q_i - Q_1^* e^{-\phi \|I_i - I_1^*\|^2}$ 。距离聚类中心越近的样本,其 Q_i 值就越小,成为下一个聚类中心的概率也就越低。

步骤4 迭代执行上述过程,直到第 t 次迭代的 Q_i^* 满足 $|Q_{t+1}^* - Q_t^*| < \varepsilon$ 为止。其中, ε 为一个很小的正数,表示第 $t+1$ 次迭代和前一次的最大 Q 值几乎无变化。根据相关研究^[12],通常取 $\varepsilon = 10^{-3}$ 。

4.3 初始规则库构建

首先将CIDD数据库中的变量值规范化到区间 $[0,1]$,然后通过提出的改进型FCM对训练样本进行聚类,以此获得相应的模糊规则。

由于所有模糊规则具有相同的变量集,所以可将其表示为 $v_1, v_2, \dots, v_N$ 。然后从每个集群 C_i 中提取II型模糊规则 i ,其形式为^[13]:

If v_1 为 $\tilde{A}_{i1}$, v_2 为 $\tilde{A}_{i2}$, ..., v_N 为 $\tilde{A}_{iN}$

Then op 为 $c_i = \omega_{i0} + \omega_{i1}v_1 + \dots + \omega_{iN}v_N$

其中, $\tilde{A}_{ij}$ 为对应第 i 个规则的变量 v_j 的II型模糊集, $1 \leq j \leq N$, op 为5个输出类之一, ω_{ij} 为与第 i 个规则的变量 v_j 相关联的真实权重。本文对 $\tilde{A}_{ij}$ 采用高斯隶属度函数,定义为 $\mu_{\tilde{A}_{ij}}(v_j) = \text{gauss}(u; \text{gauss}(v_j; m_{ij}^l; \sigma_{ij}^l))$ 且 $u \in [0,1]$ 。每个模糊规则都将均值(m)和偏差(σ)作为其前件参数,将规则权值(ω)作为其后件参数。这个规则权值和前件参数的值用以计算规则的贴近程度。

4.4 规则库优化

在通过聚类技术生产的规则中可能存在冗余,且规则的准确性不高。为此,本文通过细化所涉及的前件和后件参数提高规则的精确度。

首先,将 m 和 σ 分别单独设定为常数来优化



后件参数, 然后, 将 ω 值保持为常数来调优该规则的 m 和 σ 。重复这种细化过程直到产生近似最优精度, 即 $\text{MSE} < \text{MSE}_{\text{predefined}}$ 。对于形式为 $[I_1, O_1], [I_2, O_2], \dots, [I_K, O_K]$ 的 K 个训练样本, 这个优化阶段的目标是最小化目标函数, 即系统输出的均方误差 (mean square error, MSE) 值, 如式 (9) 所示:

$$E_j = \frac{1}{2} [y(I_j) - O_j]^2, 1 \leq j \leq K \quad (9)$$

其中, $y(I_j)$ 为实际输出, O_j 为第 j 个样本所期望的输出。

在调优规则的 m 时, 将第 i 个规则的规则权重 ω 值保持为常数, 其表达式如式 (10) 所示。偏差 σ 的调优表达式与此类似。

$$m_{ij}(l+1) = m_{ij}(l) - \left(\frac{\alpha(y(I_l) - O_l)(I_{il} - m_{ij})F(m_{ij}, \sigma_{ij}; I_{il})}{2\sigma_{ij}^2} \times \frac{\left(\prod_{q=1, q \neq j}^{N_{R_i}} \bar{u}_{\tilde{A}_{iq}} \right) (\omega_{iq} - y(I_j))}{\sum_{k=1}^L g^{-k} + \sum_{k=L+1}^M g^{-k}} \right) \quad (10)$$

在调优 ω 时, 通过保持均值和偏差为定值, 其表达式如式 (11) 所示。

如果 $j \leq L$:

$$\omega_{ij}(l+1) = \omega_{ij}(l) - \left\{ \frac{\alpha}{2} \frac{(y(I_l) - O_l) \times \prod_{q=1}^{N_{R_i}} \bar{u}_{\tilde{A}_{iq}}}{\sum_{k=1}^L \bar{g}^i + \sum_{k=L+1}^M \underline{g}^i} \right\} \quad (11)$$

如果 $j > L$:

$$\omega_{ij}(l+1) = \omega_{ij}(l) - \left\{ \frac{\alpha}{2} \frac{(y(I_l) - O_l) \times \prod_{q=1}^{N_{R_i}} \underline{u}_{\tilde{A}_{iq}}}{\sum_{k=1}^L \bar{g}^i + \sum_{k=L+1}^M \underline{g}^i} \right\} \quad (12)$$

其中, α 为学习率参数, N_{R_i} 为输入第 i 个规则的数量。 $\underline{u}_{\tilde{A}_{iq}}$ 为模糊集 $\tilde{A}^{iq} - q^i$ 的下限隶属度值,

$\bar{u}_{\tilde{A}_{iq}}$ 为模糊集 $\tilde{A}^{iq} - q^i$ 的上限隶属度值, L 和 M 值可根据迭代的 Karnik-Mendel 方法^[14]来导出。另外, F 、 $\bar{g}^k$ 和 $\underline{g}^k$ 分别定义如下:

$$F(m_{ij}, \sigma_{ij}; I_{ij}) = \exp \left(-\frac{1}{2} \left(\frac{I_{il} - m_{ij}}{\sigma_{ij}^2} \right)^2 \right) \quad (13)$$

$$\bar{g}^k = \prod_{q=1}^{N_{R_i}} \bar{u}_{\tilde{A}_{iq}}(I_q) \quad (14)$$

$$\underline{g}^k = \prod_{q=1}^{N_{R_i}} \underline{u}_{\tilde{A}_{iq}}(I_q) \quad (15)$$

这个阶段最终构建一个紧凑的模糊规则库。

当云用户请求访问云服务时, 该系统会将用户的连接样本作为模糊推理的输入, 通过模糊规则推理出该用户是否为入侵者。

4.5 基于规则库的入侵分类

获得模糊规则库后, 当有网络连接样本输入时, 通过模糊推理给出的隶属度值计算样本与各类的相似度, 以此来分类样本, 实现入侵检测。其中, 相似度由样本各语言变量的隶属度表示。假设, 输入样本特征为 $I = [I_1, I_2, \dots, I_N]$, 那么其与第 j 类的相似度表达式为:

$$T_j = \prod_{i=1}^N d_{i,j} = \max_{1 \leq k_{i,j} \leq L_{i,j}} [u_{i,j}(I_i)] \quad (16)$$

其中, $u_{i,j}(I_i)$ 表示 I_i 对应于模糊规则 $\tilde{A}_{i,j}$ 的隶属度。最后, 将与其具有最高相似度的类作为该样本的类别。

5 实验及分析

5.1 实验设置

在 Intel 酷睿 i5 处理器、2.5 GHz 主频、8 GB 内存和 Windows 7 平台上, 通过 MATLAB 实现提出的入侵检测方法。由于条件有限, 无法在实际云平台上进行实验。为此文本选择了一个云入侵检测数据集^[15]进行仿真实验, 其由用户远程访问云平台时可能存在的攻击数据和正常访问数据组成。

表1 特征降维后的特征集

特征名	描述
service	在目标主机的网络服务
dst_bytes	从目标主机到源主机的数据流量
count	2 s 内与当前连接具有相同目标主机的连接数
serror_rate	2 s 内, 在与当前连接具有相同目标主机的连接中, 出现“SYN”错误的连接的百分比
srv_diff_host_rate	2 s 内, 在与当前连接具有相同服务的连接中, 与当前连接具有不同目标主机的连接的百分比
dst_host_srv_count	前 100 个连接中, 与当前连接具有相同目标主机相同服务的连接数
src_bytes	源主机到目标主机的数据流量
logged_in	登录成功
srv_count	2 s 内与当前连接具有相同服务的连接数
srv_rerrAor_rate	2 s 内, 在与当前连接具有相同服务的连接中, 出现“REJ”错误的连接的百分比
dst_host_count	前 100 个连接中, 与当前连接具有相同目标主机的连接数
dst_host_diff_srv_rate	前 100 个连接中, 与当前连接具有相同目标主机不同服务的连接所占的百分比

CIDD 数据集中共包含四大类攻击, 即拒绝服务(denial of service, DoS)攻击、探测(probe)攻击、远程到本地(remote to login, R2L)攻击和非授权访问(user to root, U2R)攻击。每个连接记录具有 41 个特征, 用于描述总共 24 种子攻击类型。原始数据集大小为 744 MB, 拥有 4 940 000 条记录。

随机选择了 20 000 条记录作为实验数据集, 其中, 12 000 条为正常数据, 8 000 条为入侵数据, 入侵数据包含了 24 种子攻击类型, 且每种攻击的数据记录不少于 100 条。将其中的 60%数据作为训练集, 剩下的 40%数据作为测试集。基于训练集获得模糊规则, 并存储在模糊规则库中。在测试阶段, 对于从用户发送到云的服务请求数据, 基于模糊规则库的推理决定是否允许该请求。

5.2 特征选择

若使用网络连接数据中的所有 41 个特征构建 IDS 模型, 会大大增加计算量。此外, 数据集中包含不相关的特征, 则会影响入侵检测的准确性。因此, 需要通过特征选择算法降低特征维度。本文使用 MIFS 算法将原始特征数量减少到 12 个, 见表 1。

另外, 数据集中的特征具有不同的数据形式, 因此需要对特征进行预处理, 将其规范化到[0,1]。

5.3 验证实验

首先, 在整个训练集和测试集上对提出的方

法进行验证性实验, 入侵分类的混淆矩阵见表 2 和表 3 (其中 normal 指正常类), 其中所列结果为 3 次实验的平均值。可以看出, 提出的方法在训练集上的平均检测率达到了 98.93%, 在测试集上达到了 98.48%。其中, 平均检测率为 5 种攻击类别正确分类率的平均值。由于检测模型是根据训练集训练获得, 所以在训练集上的检测率略微较高。实验结果证明了提出方法的有效性。

表2 训练集上分类的混淆矩阵

类别	分类率				
	normal	probe	DoS	U2R	R2L
normal	98.31%	0.19%	0.76%	0.25%	0.49%
probe	0.13%	99.47%	0.05%	0.28%	0.07%
DoS	0.87%	0.03%	98.94%	0.1%	0.06%
U2R	0	0.25%	0.41%	98.92%	0.42%
R2L	0	0.4%	0.13%	0.46%	99.01%

此外, 为了验证提出方法在不同样本数量下的检测性能。选择数量为 5 000~20 000 的样本集, 其中各样本集里都包含与完整样本集中同比例的各种攻击样本, 同样将其中 60%作为训练集, 其他 40%作为测试集。平均检测率见表 4。可以看出, 在样本数量较小时, 检测率较低, 随着样本数量的增加, 检测率有所提高并趋于稳定。



表3 测试集上分类的混淆矩阵

类别	分类率				
	normal	probe	DoS	U2R	R2L
normal	98.33%	0.12%	0.52%	0.32%	0.71%
probe	0.18%	98.52%	0.25%	0.68%	0.37%
DoS	1.17%	0.33%	97.82%	0.52%	0.16%
U2R	0.03%	0.24%	0.61%	98.73%	0.39%
R2L	0.07%	0.31%	0.16%	0.45%	99.01%

表4 不同大小样本集下的平均检测率

样本数量/个	检测率	
	训练集	测试集
5 000	95.69%	94.64%
8 000	97.10%	96.87%
11 000	97.70%	97.26%
14 000	98.67%	98.35%
17 000	98.94%	98.43%
20 000	98.93%	98.48%

5.4 比较实验

为了进一步证明提出方法的优越性,将其与现有基于聚类技术构建模糊规则集的入侵检测方法进行比较。分别为基于传统 K 均值聚类、参考文献[7]提出的改进 K 均值聚类和参考文献[8]提出的模糊 C 均值聚类的入侵检测方法。为了公平比较,在训练和测试过程中,都采用 MIFS 进行特征选择。在样本集数量为 5 000~20 000 的条件下,各种方法在训练集和测试集上的检测率如图 2 和图 3 所示。

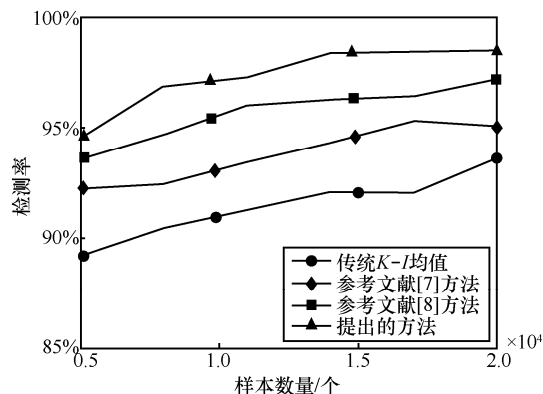


图2 各种方法在训练集上的检测率比较

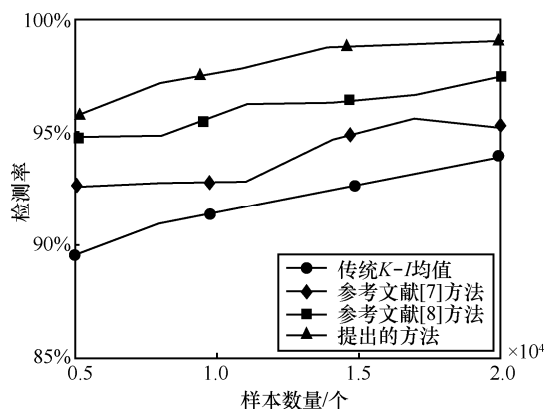


图3 各种方法在测试集上的检测率比较

可以看出,在不同样本数量下,提出方法都获得了最高的检测率。这是因为,传统 K 均值聚类算法对初始聚类中心敏感,致使其聚类效果不好,从而在此基础上获得的模糊规则准确性也较差。参考文献[7]提出的改进型 K 均值聚类算法在一定程度上解决了初始值问题,所以其性能比传统 K 均值聚类优越。然而, K 均值聚类是一种硬性聚类方法,由于在实际应用中,某些对象不是绝对的只属于某一集群,还可能属于其他集群,某个对象与某个集群的关系是模糊的。所以基于 K 均值聚类的方法没有参考文献[8]采用的模糊 C 均值聚类的性能好。本文方法同样采用了模糊 C 均值聚类,但融入了聚类中心相似度量因子,解决了其初始聚类中心选择问题,所以能够提高其聚类性能。另外,本文方法还对通过聚类获得的模糊规则集进行了优化,构建了一个更为准确的规则集,所以获得了最佳的入侵检测性能。

6 结束语

本文提出一种基于 IFCM 和模糊推理的网络入侵检测方法。利用提出的 IFCM 聚类方法对训练样本集进行聚类,获得初始模糊规则库。通过对每个规则的前件参数和后件参数进行调优以此优化规则库。基于规则库对输入连接数据进行入侵检测。在云入侵检测数据集上,将提出的方法与现有的几种聚类方法进行了比较,结果表明,

提出方法在检测率方面具有优越性。

参考文献:

- [1] 罗亮, 吴文峻, 张飞. 面向云计算数据中心的能耗建模方法[J]. 软件学报, 2014, 25(7): 1371-1387.
LUO L, WU W J, ZHANG F. Energy modeling based on cloud data center[J]. Journal of Software, 2014, 25(7): 1371-1387.
- [2] FOSSACECA J M, MAZZUCHI T A, SARKANI S. MARK-ELM: application of a novel multiple kernel learning framework for improving the robustness of network intrusion detection[J]. Expert Systems with Applications, 2015, 42(8): 4062-4080.
- [3] 石云, 陈钟, 孙兵. 基于均值聚类分析和多层核心集凝聚算法相融合的网络入侵检测[J]. 计算机应用研究, 2016, 32(2): 518-520.
SHI Y, CHEN Z, SUN B. Networks intrusion detection based on integrating k-mean clustering analysis and multilayer core-set agglomerative algorithm[J]. Computer Applications and Software, 2016, 32(2): 518-520.
- [4] 陆婷婷, 韩旭. 面向 MANET 报文丢弃攻击的模糊入侵检测系统[J]. 电信科学, 2016, 32(10): 124-129.
LU T T, HAN X. Fuzzy intrusion detection system for MANET packet dropping attack[J]. Telecommunications Science, 2016, 32(10): 124-129.
- [5] SHAMSHIRBAND S, AMINI A, ANUAR N B, et al. D-FICCA: a density-based fuzzy imperialist competitive clustering algorithm for intrusion detection in wireless sensor networks[J]. Measurement, 2014, 55(9): 212-226.
- [6] 李洪成, 吴晓平, 陈燕. MapReduce 框架下支持差分隐私保护的 K-means 聚类方法[J]. 通信学报, 2016, 37(2): 124-130.
LI H C, WU X P, CHEN Y. K-means clustering method preserving differential privacy in MapReduce framework[J]. Journal on Communications, 2016, 37(2): 124-130.
- [7] TIAN L, JIANWEN W. Research on network intrusion detection system based on improved K-means clustering algorithm[C]//International Forum on Computer Science-Technology and Applications, Dec 25-27, 2009, Chongqing, China. Piscataway: IEEE Press, 2009: 76-79.
- [8] XIE L, WANG Y, CHEN L, et al. An anomaly detection method based on fuzzy C-means clustering algorithm[J]. Proceedings of the International Symposium on Networking & Netwo, 2014, 24(5): 56-63.
- [9] 魏莎莎, 陆慧娟, 金伟, 等. 基于云平台的互信息最大化特征提取方法研究[J]. 电信科学, 2013, 29(10): 38-42.
WEI S S, LU H J, JIN W, et al. Maximum mutual information feature extraction method based on the cloud platform[J]. Telecommunications Science, 2013, 29(10): 38-42.
- [10] 徐峻岭, 周毓明, 陈林, 等. 基于互信息的无监督特征选择[J]. 计算机研究与发展, 2012, 49(2): 372-382.
XU J L, ZHOU Y M, CHEN L, et al. An unsupervised feature selection approach based on mutual information[J]. Journal of Computer Research and Development, 2012, 49(2): 372-382.
- [11] AMIRI F, REZAEI YOUSEFI M, LUCAS C, et al. Mutual information-based feature selection for intrusion detection systems[J]. Journal of Network & Computer Applications, 2011, 34(4): 1184-1199.
- [12] SONG J, ZHU Z, SCULLY P, et al. Selecting features for anomaly intrusion detection: a novel method using fuzzy C means and decision tree classification[J]. Cyberspace Safety and Security, 2013, 8(3): 299-307.
- [13] 据春华, 鲍福光, 戴俊彦. 一种融入公众情感投入分析的微博话题发现与细分方法[J]. 电信科学, 2016, 32(7): 97-105.
JU C H, BAO F G, DAI J Y. Discovery and segmentation method in micro-blog topics based on public emotional engagement analysis[J]. Telecommunications Science, 2016, 32(7): 97-105.
- [14] LIU X, QIN Y, WU L. Fast and direct Karnik-Mendel algorithm computation for the centroid of an interval type-2 fuzzy set[C]//IEEE International Conference on Fuzzy Systems, June 10-15, 2012, Brisbane, QLD, Australia. Piscataway: IEEE Press, 2012: 1-8.
- [15] KHOLIDY H A, BAIARDI F. CIDD: a cloud intrusion detection dataset for cloud computing and masquerade attacks[C]//Ninth International Conference on Information Technology-New Generations, April 11-18, 2012, Las Vegas, NV, USA. Piscataway: IEEE Press, 2012: 397-402.

[作者简介]



张春琴(1977-), 女, 浙江工业职业技术学院副教授, 浙江工业大学访问学者, 主要从事网络安全、云计算方面的研究工作。



谢立春(1974-), 男, 浙江工业职业技术学院副教授, 入选浙江省“151人才工程”, 主要从事网络安全方面的研究工作。