DOI:10.20033/j.1003-7241.(2026)01-0176-05

# 基于半监督聚类的医院网络防入侵自动检测方法

郑晓渊, 孙 婷

(常州市第二人民医院,江苏 常州 213000)

**摘 要**:为了应对日益复杂的医院网络安全问题,提高医院信息系统安全水平。研究基于深度学习技术与半监督聚类设计了入侵自动检测模型,利用改进的自编码器建立了自编码降维模型;并借助半监督聚类实现入侵识别。实验结果表明,研究设计的改进策略有效提升了自编码的解码能力,重构结果与真实数据最为接近。改进对半监督聚类的兰德系数为0.907、纯度为0.869,标准互信息为0.837,聚集度和分离度良好。入侵检测模型检测率高于90.0%,误报率低于15%。研究设计的入侵自动检测模型对于提高医院信息系统安全,保护医院和患者的利益具有重要意义。

**关键词**:半监督聚类;信息化;入侵检测;医院;网络安全;深度学习

**中图分类号**: TP393.081 **文献标志码**: A **文章编号**: 1003-7241(2026)01-0176-05

# Automatic detection method for hospital network intrusion prevention based on semi-supervised clustering

ZHENG Xiaoyuan, SUN Ting

( Changzhou No. 2 People's Hospital, Changzhou 213000, Jiangsu, China )

**Abstract**: In order to cope with the increasingly complex hospital network security problems and improve the security level of hospital information system. The study designs an intrusion automatic detection model based on deep learning technology and semi-supervised clustering, establishes a self-coding dimensionality reduction model using an improved self-coder; and realizes intrusion recognition with the help of semi-supervised clustering. The experimental results show that the improved strategy designed by the study effectively enhances the decoding ability of self-coding, and the reconstruction results are closest to the real data. The improved pair of semi-supervised clustering has a rand index of 0.907, a purity of 0.869, a standard mutual information of 0.837, and good aggregation and separation. The intrusion detection model has a detection rate higher than 90.0% and a false alarm rate lower than 15%. The intrusion automatic detection model designed in this study is important for improving the security of hospital information system, promoting hospitals, and protecting the interests of hospitals and patients.

**Keywords**: clustering; information; intrusion detection; hospital; network security; deep learning

随着互联网产业的飞速发展,医院信息化的建设逐渐深入,信息系统在医院日常运行中的作用日益凸显。医院信息网络系统的主要目的是提高医疗信息化管理效率,实现医院的日常运营、患者病历信息管理、医疗资源调配以及医学研究等多种功能[1]。医院信息系统涉及大量敏感信息的传输和存储,但海量的网络接入点、广泛交互的物联网给医院网络安全带来了巨大的挑战。各种恶意攻击行为可能导致患者信息泄露、身份盗窃、数据丢失或损坏等严重后果,影响医院业务连续性和患者服务质量[2]。网络安全对医院的信息化运行起到了决定性作用,医院网络系统需要配备高级的危险防御机制。目前,医院网络中心常通过各种物理防护手段、防护墙以及入侵报警系统维护网络安全,同时入侵检测技术也取得了一定进展。为了防止数据污染,刘广睿等利用模糊测试改进了生成对抗网络,基于生产的边缘样本设计了支持污染过滤的智能网络入侵检测模型,该模型检测率平均提升12.50%[3]。Hazman C等展开了物联网入侵检测系统的研究分析,结合AdaBoost和特征选择技术设计了一种异常检测模型,经公开数据集验证,该方法在召回率与精度方面具有良好的性能[4]。但现有入侵检测模型多为浅层模型,面对复杂的入侵行为,检测精度较低。为了提高医院网络应对网络安全风险的技术能力,研究利用性能优异的深度学习与半监督聚类技术设计了医院网络防入侵自动检测模型。研究的创新性主要体现在两个方面,其一研究提出了自编码器与半监督的改进策

**收稿日期**:2024-03-27;**录用日期**:2024-04-16

**基金项目**:江苏省自然科学基金资助项目(BK20190571)

**作者简介**:郑晓渊(1985—),男,本科,工程师,研究方向:医院信息化。

**通信作者**:孙 婷(1986—),女,本科,工程师,研究方向:医院信息化。

**引用本文**:郑晓渊, 孙婷. 基于半监督聚类的医院网络防入侵自动检测方法[J]. 自动化技术与应用, 2026,45(1):176-179,184. (ZHENG Xiaoyuan, SUN Ting. Automatic detection method for hospital network intrusion prevention based on semi-supervised clustering[J]. Techniques of Automation and Applications, 2026,45(1):176-179,184.)

略,丰富了深度学习与半监督聚类的理论研究;其二,研究引入半监督聚类到入侵检测中,有望解决入侵模型过度依赖标记数据的困境。

## 1　医院网络防入侵自动检测技术研究

### 1.1　自编码网络降维模型设计

随着各种网络点的接入、网络模式的发展,入侵数据、网络攻击行为逐渐变得复杂多样,传统的浅层入侵检测模型难以应对特征复杂且高维的数据,而高维数据也可能造成严重的检测误差与“维数灾难”,因此入侵检测模型的设计前提是进行数据降维模型的研究与分析,以提高入侵检测的效率与性能[5]。数据降维是指通过技术手段实现高维数据到低维空间的处理与分析,可消除冗余数据,减少数据的维数,提高数据处理和分析的效率。传统的降维方法面对海量高维数据存在一些不足,如时间与空间的复杂度过高、存储需求过大、稀疏数据与非线性关系的低适用性等[6-7]。深度学习具有较强的自动学习能力,可自动从海量数据中提取有用的特征与模式,多层次的网络结构使得深度学习具备逐层提取数据的抽象特征的能力。因此,研究选择利用深度学习的方式设计数据降维模型[8]。

深度自动编码器(deep auto-encoder, DAE)由多个自编码器堆叠而成,DAE 包含较多的隐藏层,可捕捉高级别的特征和复杂模式[9]。DAE 通过最小化输入数据与重构数据之间的差异来学习特征表示,其全连接隐藏层为受限玻尔兹曼机(restricted boltzmann machine, RBM), DAE 和 RBM 的结构如图 1 所示。

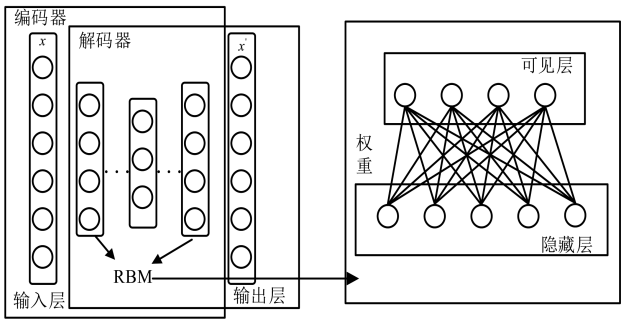



**图 1　深度自动编码器和隐藏层结构示意图**

Fig. 1　Schematic diagram of the structure of the depth autoencoder and the hidden layer

由图 1 可见,DAE 主要结构包含一个编码器和一个解码器,编码器将输入数据压缩为较低维度的表示,解码器将该低维度表示恢复为与原始输入数据尺寸相同的重构。RBM 的可见层 $v$ 与隐藏层 $h$ 的能量取值 $E(v,h)$ 计算过程为

$$E(v,h) = -\sum_{i=1}^{I} \boldsymbol{c}_i v_i - \sum_{j=1}^{J} \boldsymbol{b}_j h_j - \sum_{j=1}^{J}\sum_{i=1}^{I} W_{ji} v_i \boldsymbol{b}_j \tag{1}$$

式中,$\boldsymbol{c}$、$\boldsymbol{b}$ 分别表示 $v$、$h$ 的偏置向量,且 $\boldsymbol{c} \in \mathbb{R}^I$、$\boldsymbol{b} \in \mathbb{R}^J$;$W$ 表示 $v$ 和 $h$ 之间的连接权重,$W \in \mathbb{R}^{J\times I}$。研究采用 Gibbs 测度法计算可见层向量的出现概率 $p(v)$,计算过程为

$$p(v) = \sum_h p(v,h) = \frac{\sum_h e^{-E(v,h)}}{\sum_{v,h} e^{-E(v,h)}} \tag{2}$$

可见层、隐藏层神经元被激活的概率为

$$\begin{cases} p(h_j = 1 \mid v) = \sigma\left(\boldsymbol{b}_j + \sum_{i=1}^{I} W_{ji} v_i\right) \\ p(v_j = 1 \mid h) = \sigma\left(\boldsymbol{c}_j + \sum_{i=1}^{I} W_{ji} h_i\right) \end{cases} \tag{3}$$

经过 RBM 的堆叠与展开,得到 DAE 模型。将 RBM 的权重作为编码器全连接层的权重初值;解码器对称构造,解码器权重为编码器权重的转置,DAE 模型训练过程如图 2 所示[10-11]。

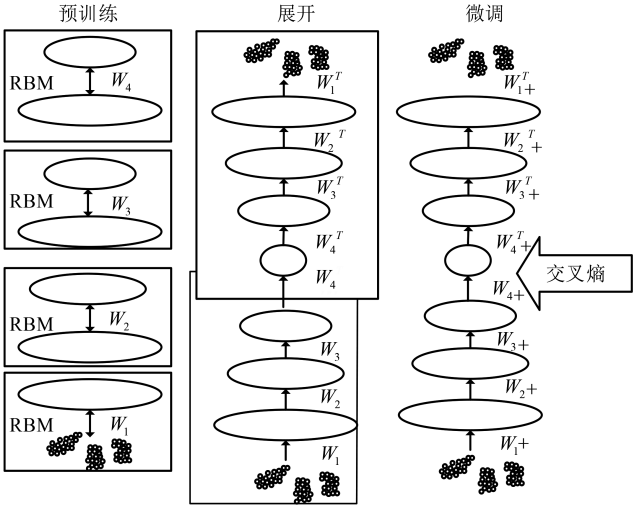



**图 2　DAE 训练过程示意图**

Fig. 2　Schematic diagram of the DAE training process

编码过程与解码过程见式(4)。

$$\begin{cases} Z^{(2)} = W^{(1)} x + b^{(1)} \\ a^{(2)} = f(Z^{(2)}) \\ Z^{(3)} = W^{(2)} a^{(2)} + b^{(2)} \\ a^{(3)} = g(Z^{(3)}) \end{cases} \tag{4}$$

式中,$a^{(2)}$、$a^{(3)}$ 表示编码、解码的结果;$f$、$g$ 表示激活函数;$x$ 表示输入数据。$b^{(1)}$ 是编码器全连接层的偏置项,对应于从输入层到第一个隐藏层的线性变换;$b^{(2)}$ 是解码器全连接层的偏置项,对应于从隐藏层到输出层的线性变换。DAE 通过最小化输入数据与重构数据之间的差异来学习特征表示,这限制了 DAE 重构数据的准确性,为了便于半监督聚类模型的检测,研究在传统 DAE 模型中加入了惩罚项,达到提高输入数据与重构数据之间的相似性[12]。损失函数表达式为

$$\boldsymbol{J}_{Rm}(W,b) = \boldsymbol{J}_{\text{sparse}}(W,b) + \lambda(\|\boldsymbol{A}x - \boldsymbol{A}a^{(3)}\|^2) \tag{5}$$

式中,$\boldsymbol{A}$ 表示特征关系矩阵,其中 $\boldsymbol{A}x$ 表示输入数据特征关系矩阵,$\boldsymbol{A}a^{(3)}$ 表示输出数据的特征关系矩阵;$\boldsymbol{J}_{\text{sparse}}(W,b)$ 表示加入了稀疏性和系数正则项 $\lambda$ 的损失函数。此外,研究将目标函数确定为重构误差函数。权重对深度学习模型的加权影响较大,研究引入 Xavier 初始化方法对权重进行初始化,避免模型陷入局部最优的问题。Xavier 计算

表达式为

$$W \sim U\left[-\frac{\sqrt{6}}{\sqrt{n_i+n_{i+1}}},\frac{\sqrt{6}}{\sqrt{n_i+n_{i+1}}}\right] \tag{6}$$

式中，$n_i$、$n_{i+1}$ 分别表示输入、输出单元。最后，为了使 DAE 学习未被噪声污染的数据特征，研究在输入的数据中随机加入了高斯白噪声，并随机毁坏部分数据增加模型的鲁棒性。

### 1.2 入侵检测模型设计

完成高维数据降维处理后，研究采用半监督聚类算法进行入侵检测模型设计。入侵检测模型的设计有利于及早发现信息系统的漏洞与风险，可有效防止医疗数据与隐私的泄漏，医院应当进行有效的入侵检测，采取相应的安全措施，以保护患者隐私和医院网络的安全[13]。半监督聚类算法结合了无监督聚类和有监督分类技术，通过无监督聚类算法聚类未标记样本利用一小部分标记样本的信息来对聚类结果进行修正和调整，可实现少量标记样本的有效数据聚类。半监督聚类提高了标记数据较少时的聚类准确性，降低了人工标注的成本[14-15]。

基于密度的聚类算法（density-based spatial clustering of applications with noise，DBSCAN）将数据点划分为若干个密度相连的簇，簇内的数据点与簇外的数据点有明显的密度差异。DBSCAN 设定阈值 Minpts、r 表示样本邻域半径，若该点的邻域内的点个数大于等于阈值，则标记为核心对象；若该点的邻域内的点个数不足，则将该点标记为噪声点；当若该点的邻域内的点个数不足，但落在某个核心点的邻域内，该点记为边界点。DBSCAN 聚类过程如图 3 所示。

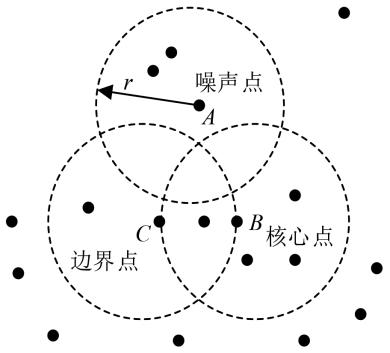



**图 3 DBSCAN 聚类过程示意图**

Fig. 3 Schematic diagram of the DBSCAN clustering process

由图 3 可见，相比其他基于距离的聚类算法，DBSCAN 可发现任意形状的空间聚类，聚类速度较快，且能高效地处理噪声点，不易受异常点干扰聚类效果。但 DBSCAN 对内存的消耗较大，且过多依赖于参数的选择，当空间聚类的密度不均匀、聚类间距差相差很大时，聚类效果较差。对此，研究对传统的 DBSCAN 聚类算法进行了改进，设计了一种基于密度的多中心半监督聚类（density-based multi center semi supervised clustering，DBMCSSC）算法。该方法首先利用距离测度进行数据划分，在数据划分区利用 DBSCAN 进行聚类，局部聚类结果合并后得到整体聚类结果。

首先进行参数选择，研究使用自适应的参数选择方法确定 DBSCAN 的阈值 Minpts 与半径 $r$。定义 $p$ 点到 $k$ 个邻近点的距离，记为 $k$-dis，统计分析 $k$-dis 取值分布情况。根据 $k$-dis 分布曲线的拐点和 $r$ 邻域内的点个数确定阈值 Minpts 与半径 $r$。为了减少 DBSCAN 邻域查询的冗余操作，研究采取的策略是仅遍历部分对象扩展类，并在噪声数据处理过程中对处理与未处理的点进行分类。选择的遍历对象需满足距离核心对象最远和遍历对象间距最大的条件。

研究对数据的划分采用距离相似性测度法。定义数据集为 $X=\{x_1,x_2,\cdots,x_n\}$，假设簇类数为 $k$，簇的中心向量表示为 $\mu^{(1)},\mu^{(2)},\cdots,\mu^{(k)}$。分别计算数据 $x_i$ 与簇中心向量的距离，按照距离远近将 $x_i$ 划分到簇中心对应的数据集 $C$ 中。数据划分可一定程度缓解密度不均匀、聚类间距差相差很大带来的聚类误差。按照划分区域进行局部聚类，利用 DBSCAN 的鲁棒性提高聚类精度。最后标记数据的类型，将类型相同的数据集合并。

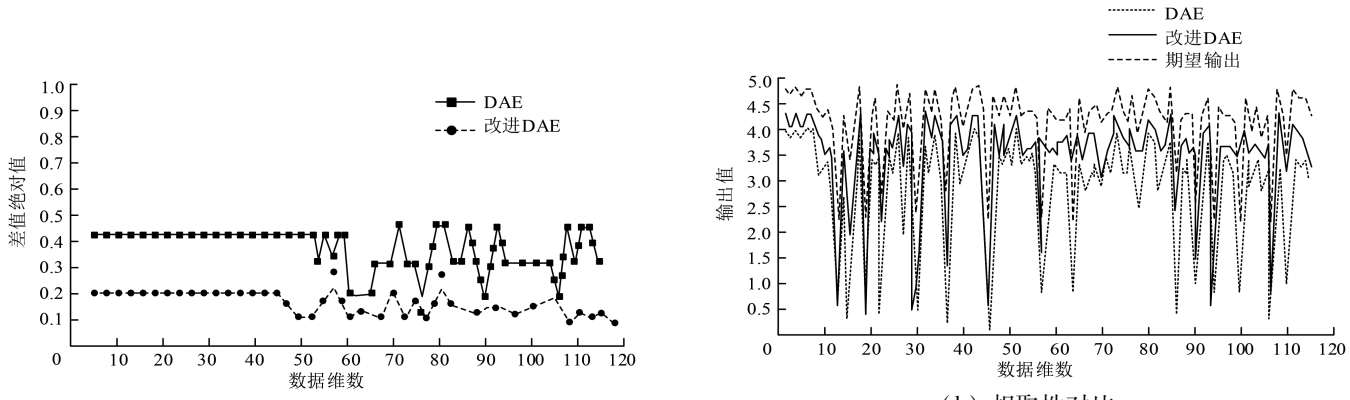
## 2 模型性能测试与效果分析

为了验证研究设计的入侵检测模型的性能，研究进行了测试实验分析。实验环境的操作系统选用 Linux，深度学习框架使用 Pytorch 1.11，编程语言为 Python 3.8，图形处理器为 GeForce RTX 2080ti，中央处理器为 Inter i5。实验数据集为 KDD'99、IoT-23、BoT-IoT 和 Edge-IIoT 数据集。KDD'99 记录了十种不同类型的网络入侵和正常网络流量数据，该数据集包含大量的噪声和异常行为。IoT-23 包含了 23 种不同类型物联网设备遭受不同类型攻击时的网络流量；BoT- IoT 可用于各类数据攻击和泄漏；Edge-IIoT 涵盖了工业控制系统在正常运行和遭受多种攻击时的网络流量数据。

首先进行 120 维数据到 5 维数据的降维实验，任意选取 KDD'99 数据集中的一段数据进行分析。对比分析 DAE 模型改进前后实际输出与期望输出的差值以及重构数据和原始数据的相似性，实验结果如图 4 所示。由图 4(a)可见，研究改进之后的 DAE 模型的差值小于传统的 DAE 模型，绝对值在 0.0-0.2 区间内小范围波动。传统的 DAE 模型差值绝对值波动范围较大，最大波动区间为 0.32，表明改进 DAE 模型的重构数据更接近期望值。结合图 4(b)分析，改进之后 DAE 模型的重构数据与原始数据的局部相似性更大，在维数变化过程中更接近原始数据。综合而言，研究设计的改进策略提高了 DAE 模型的解码能力，有助于数据的降维分析，有利于入侵检测的性能 0 提升。

分析研究设计的半监督聚类算法的聚类性能，聚类效果直接影响入侵检测的准确性。纯度、选择标准互信息（normalized mutual information，NMI）、兰德系数（rand index，RI）的统计结果如表 1 所示。纯度计算聚类结果中被正确分类的样本数量与总样本数量之间的比例；NMI 衡量聚类结果与真实标签之间的相似度；RI 衡量聚类结果中符合真实标签的配对数与总配对数之间的比例，三种评价指标取值越大，聚

类模型的聚类结果越好,聚类结果与真实标签的一致性越高。由表 1 可见,研究改进的 DBMCSSC 模型的纯度、NMI 以及 RI 表现较优,在 IoT-23 数据集上最大 RI 为 0.907、纯度为 0.869,Edge- IIoT 数据集上最大 NMI 为 0.837。同等实验环境下,优于其他两种聚类模型,研究选择的密度聚类方法以及改进策略均起到了优化聚类效果的成绩。




(a) 差值对比 (b) 相取性对比

**图 4 DAE 模型改进前后性能对比**

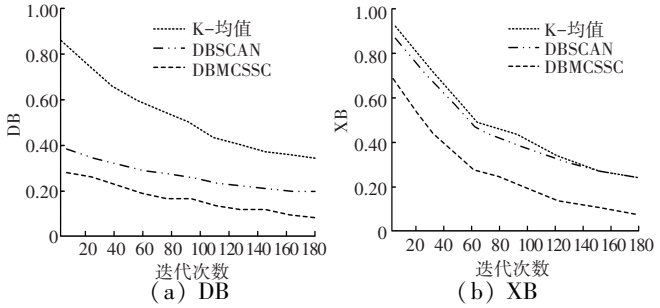
Fig. 4 Performance comparison of the DAE model before and after improvement

**表 1 不同聚类模型的聚类效果评价**

Tab. 1 Evaluation of clustering effects of different clustering models

| 聚类模型 | K-均值 | | | DBSCAN | | | DBMCSSC | | |
|---|---|---|---|---|---|---|---|---|---|
| | 纯度 | NMI | RI | 纯度 | NMI | RI | 纯度 | NMI | RI |
| KDD'99 | 0.646 | 0.627 | 0.702 | 0.742 | 0.799 | 0.762 | 0.801 | 0.813 | 0.845 |
| IoT-23 | 0.694 | 0.756 | 0.692 | 0.736 | 0.749 | 0.719 | 0.869 | 0.807 | 0.907 |
| BoT-IoT | 0.639 | 0.728 | 0.636 | 0.719 | 0.706 | 0.734 | 0.811 | 0.783 | 0.807 |
| Edge-IIoT | 0.703 | 0.691 | 0.704 | 0.748 | 0.706 | 0.788 | 0.861 | 0.837 | 0.846 |

选择谢比尼指数(xie beni, XB)和戴维森堡丁指数(davies-bouldin index, DB)指标对不同聚类模型更深入地进行分析,实验结果如图 5 所示。XB 指标衡量了聚类结果的紧密性和分离度,指标取值越小代表聚类效果越好。DB 衡量了模型的聚类结果的聚集度和分离度,DB 指标取值越小,代表簇类内距离越小和类间距离越大,越接近 0 表示聚类结果越好。由图 5(a)可见,随迭代次数的增加,不同聚类模型的 DB 值均呈现下降趋势,但 DBMCSSC 模型的 DB 值一直处于最低水平,最小 DB 值仅 0.09。由图 5(b)可见,不同聚类方法的 XB 值呈下降趋势,DBMCSSC 模型的仍表现出较为明显的取值优势,聚类结果较好。

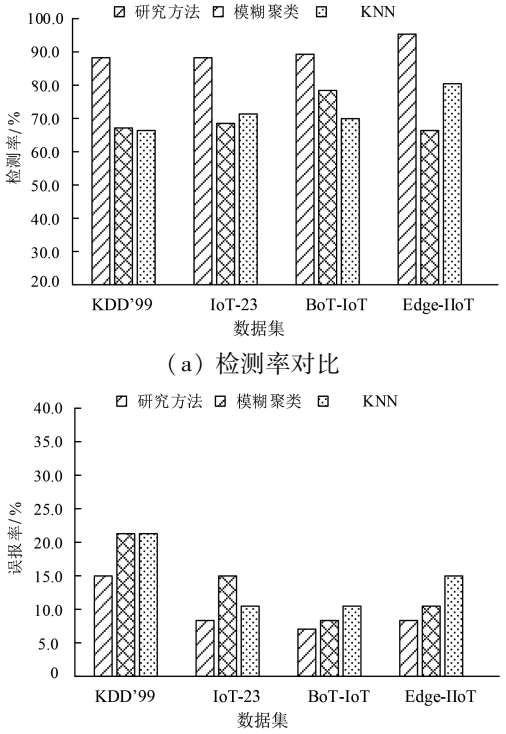



(a) DB (b) XB

**图 5 不同聚类模型评价指标统计结果对比**

Fig. 5 Comparison of statistical results of evaluation indicators of different clustering

最后,分析对比不同入侵检测模型的检测率与误报率,选择模糊聚类法和 K-近邻(k-nearest neighbor, KNN)进行性能对比,不同数据集的实验结果如图 6 所示。




(a) 检测率对比 (b) 误报率对比

**图 6 检测模型检测率与误报率对比**

Fig. 6 comparison of the detection rate and false alarm rate of the detection model

(下转第 184 页)