

基于自编码网络和聚类的入侵检测技术

周 康¹, 万 良²

(1. 贵州大学 计算机科学与技术学院, 贵州 贵阳 550025;

2. 贵州大学 软件与理论研究所, 贵州 贵阳 550025)

摘 要: 针对模糊 C 均值聚类算法的入侵检测方法易陷入局部最优, 受时间和空间复杂度约束, 检测速率低并且使用原始数据集容易陷入“维度灾难”等问题, 提出了一种基于自编码网络 (AN) 特征降维结合遗传算法 (GA) 优化模糊 C 均值算法的聚类模型 (AN-GA-FCM)。该模型采用多层限制玻尔兹曼机 (RBM) 将高维、非线性的数据双向映射到低维空间, 建立高维空间到低维空间的自编码网络, 进而使用自编码网络权值微调重构低维空间数据的最优高维表示。并利用遗传算法优化的 FCM 初始聚类中心, 避免目标函数陷入局部最优。将得到的特征降维数据集通过 GA-FCM 进行分类并在 KDD'99 数据集上进行检测, 通过与 PCA、SVM、Softmax 等传统算法的实验对比, 结果表明, 该模型具有较高的入侵检测准确率和较低分类检测时间。

关键词: 模糊 C 均值; 遗传算法; 限制玻尔兹曼机; 自编码网络; 特征降维; 双向映射

中图分类号: TP301.6

文献标识码: A

文章编号: 1673-629X(2019)05-0107-05

doi: 10.3969/j.issn.1673-629X.2019.05.023

Intrusion Detection Technology Based on Self-coded Networks and Clustering

ZHOU Kang¹, WAN Liang²

(1. School of Computer Science and Technology, Guizhou University, Guiyang 550025, China;

2. Institute of Software and Theory, Guiyang 550025, China)

Abstract: The intrusion detection method for the fuzzy C-means clustering algorithm is easy to fall into the local optimal, constrained by the time and space complexity, with low detection rate and easy to fall into the “dimensional disaster” and other problems using the original data set. For these problems, we propose a novel fuzzy C-means algorithm clustering model (AN-GA-FCM) based on genetic algorithm (GA) optimization combined with auto-encoder network (AN). This model uses multi-layer restricted Boltzmann machine (RBM) to bidirectionally map high-dimensional and nonlinear data into low-dimensional space, establishes high-dimensional space to low-dimensional autoencoder network, and then uses autoencoder network weights to fine-tune parameter, reconstructing the optimal high-dimensional representation to low-dimensional spatial data. The FCM initial clustering center optimized by the genetic algorithm is to avoid objective function falling into a local optimum. The dimensionality reduction datasets are classified by GA-FCM detected on the KDD'99 dataset. Meanwhile, compared with the traditional algorithms such as PCA, SVM and Softmax with the model, it shows that the model has higher intrusion detection accuracy and lower classification detection time.

Key words: fuzzy C-means; genetic algorithm; restricted Boltzmann machine; autoencoder network; feature dimensionality reduction; bi-directional mapping

0 引言

随着网络安全问题的层出不穷, 传统入侵检测方法已经不符合时代需要, 好的入侵检测方法应满足准确性高、较好的适应性和检测的实时性要求。对于传统的入侵检测方法, 前人做了较多研究。例如,

Denatious 等^[1]提出了利用聚类、关联规则等联合数据挖掘方法进行入侵检测; Chitrakar 等^[2]组合应用邻近聚类算法和 SVM 算法来构建异常检测模型; Srinoy 等^[3]将隶属度综合运用粗糙模糊聚类方法实现数据容量的约简, 实现特征降维等。这些方法都在入侵检

收稿日期: 2018-06-11

修回日期: 2018-10-16

网络出版时间: 2018-12-21

基金项目: 贵州省科学基金(黔科合 J 字[2011](2328) 黔科合 LH 字[2014](7634))

作者简介: 周 康(1993-) 男, 硕士研究生, 研究方向为深度学习、网络安全; 万 良, 博士后, 教授, 研究方向为信息安全。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20181221.1525.036.html>

测领域取得了突破性进展,而模糊 C 均值算法(fuzzy C-means, FCM)最早由 Bezdek^[4]于 1984 年提出,使用欧氏距离贴近的模糊隶属度矩阵方法表达分类。Wang^[5]和 Lee^[6]等将 FCM 方法应用于 IDS 中,证明了 FCM 具有良好的分类性能。

特征空间维数过多是导致 IDS 检测速率低的主要原因。黄思慧^[7]提出了一种 PCA-ELM 的入侵检测算法,在时间上和检测正确率上都有所改进;Kuang 等^[8]提出混合的 KPCA 方法与遗传算法(GA)、单独的 KPCA 方法对高维数据降维后代入分类器进行入侵识别;EDL Hoz^[9]将主成分分析(PCA)降维和自组织映射(SOM)聚类进行入侵检测,是一种常规高效的做法。

前人的研究都在输入特征维度上进行降维,提高了检测时间,但在特征维度空间呈现非线性特征时,无法较大地提高检测速度。因此文中旨在保证高分类精度的条件下,最大限度地学习到低维特征空间的高维数据表示,以提高检测速度。

文中将自编码网络(autoencoder network, AN)^[10]应用于入侵检测领域,使用遗传算法优化 FCM 算法聚类中心避免陷入局部最优,并将自编码网络降维和主成分分析降维进行比较,以及对比 FCM 和 SVM、softmax 分类算法,证明 AN-GA-FCM 入侵检测模型具有较高的检测速率和检测正确率。

1 深度学习模型

1.1 自编码网络

自编码网络通过学习具体特征向量获得数据集的压缩编码,达到降维的目的^[11]。其本质是学习一个 $h_{w,h}(x) \approx x$ 的函数,使得输出 \hat{x} 接近于输入 x 。通过学习编码器和解码器参数,使得重构误差最小,通过重构后的数据最大化表示原始输入数据。它实际上是求解目标函数稳定的过程,目标函数定义为:

$$L(f(X)) = \frac{1}{2} \sum_{i=1}^N (\hat{x}_i - x_i)^2 \quad (1)$$

其中, $f(X)$ 为 sigmoid 激活函数函数。

该网络激活方式是隐藏层 h^1 的输入由上一层 h^0 的输出 x 和它的权重 W 乘积激活。 h^2 的输入同样由 h^1 的输出 x^1 和它的权重 W^1 乘积激活,直到所有的隐藏层都被激活。

1.2 RBM 神经网络

自编码网络的核心组件之一是受限玻尔兹曼机网络(restricted Boltzmann machine, RBM),通过对自编码网络每两个隐藏层进行预训练得到高维空间映射到低维空间的矩阵表示。RBM 包含两层,可见层(visible layer)和隐藏层(hidden layer)。神经元之间

连接是层内无连接,层间全连接,其所有节点满足全概率分布 $p(v|h)$,即 Boltzmann 分布。通过使用更多的隐藏单元适应更复杂分布,增强其建模能力。RBM 的能量函数定义如下:

$$E(v|h) = -b^T v - c^T h - h^T W v \quad (2)$$

其中, W 表示连接隐藏单元和可见单元的权重; b, c 分别表示可见层和隐藏层的偏移量。

由于 RBM 的性质,即当给定可见层神经元的状态时,各隐藏层神经元的激活条件独立;反之当给定隐藏层神经元的状态时,可见层神经元的激活也条件独立,即:

$$P(v|h) = \frac{1}{Z} e^{-E(v|h)} \quad (3)$$

$$Z = \sum_{v,h} e^{-E(v|h)} \quad (4)$$

其中, Z 为归一化因子。

训练的目的是使联合概率最大,能量函数值最小。给定可见层(或隐藏层)所有神经元状态,隐藏层(或可见层)上某个神经单元被激活(取值为 1)的概率,计算 $P(h_k = 1|v)$ (或 $P(v_k = 1|h)$) 推导得到:

$$p(h|v) = \prod_i p(h_i|v) \quad (5)$$

$$p(v|h) = \prod_j p(v_j|h) \quad (6)$$

通过参数调整,从隐藏层获得的可见层 v_k 和原来的可见层 v 如果一样,那么所获得的隐藏层就是可见层的另一种表达。

由于可见单元和隐藏单元概率条件独立,样本 $p(x)$ 可以通过马尔可夫链收敛。

1.3 Gibbs 采样

在 RBM 中,使用 Gibbs 采样^[12]作为转换算子,样本包含了可见单元和隐藏单元集合,但它们是条件独立的,因此样本 $p(x)$ 可通过马尔可夫链收敛。在这个过程中,给定隐藏单元值同时对可见单元进行采样,同样给定可见单元值同时对隐藏单元采样。马尔可夫链中的一个步骤如下:

$$h^{(n+1)} \sim \text{sigm}(W^T v^{(n)} + c) \quad (7)$$

$$v^{(n+1)} \sim \text{sigm}(W^T h^{(n)} + b) \quad (8)$$

其中, $h^{(n)}$ 为马尔可夫链第 n 步所有隐藏单元集合。当 $t \rightarrow \infty$ 就可以得到样本最精确的概率值。

1.4 CD-k 算法

利用 Gibbs 采样和马尔可夫链求概率分布最大值的方法,其迭代收敛次数无法保障,难以确定步长。对比分歧算法(contrastive divergence, CD)是 Hinton 提出的快速训练 RBM 的方法,并在实践中取得了很好的效果^[13]。CD 算法的目的是希望 $p(v) \approx p_{\text{train}}(v)$,即数据分布差异性最小。两个概率分布的差异性表示为 Kullback-Leibler(KL),即:

$$\sum_{j=1}^N \text{KL}(p^0 \| h_{\theta}^{\infty}) = \sum_{j=1}^N p^0 \log \frac{p^0}{h_{\theta}^{\infty}} + (1 - p^0) \log \frac{1 - p^0}{1 - h_{\theta}^{\infty}} \quad (9)$$

$$\text{CD} = \text{KL}(p^0 \| h_{\theta}^{\infty}) - \text{KL}(p_{\theta}^n \| h_{\theta}^{\infty}) \quad (10)$$

其中, p^0 为 RBM 网络初始态的联合概率分布; p_{θ}^n 为经过 t 步迭代后的联合概率分布; h_{θ}^{∞} 为马尔科夫链末端的联合概率分布。KL 分歧作为损失函数的惩罚机制^[14], 学习的目标是为了最小化重构误差, 得到最终的损失函数形式如下:

$$\min L = \min [E(x, \hat{x}) + \sum_{j=1}^N \text{KL}(p^0 \| h_j)] \quad (11)$$

实验证明, 在 n 次迭代后, 通过梯度修正参数 θ , CD 值必将趋近于 0。

2 自编码网络的入侵检测模型构建

2.1 自编码网络特征提取模型构建

文中设计的自编码网络结构由数据输入层、中间 4 层 RBM 隐藏层和输出层组成^[15], 实现过程如下:

(1) 预训练过程。通过 RBM 预训练获得生成模型权值参数, 计算 W 和 b 的残差, 利用梯度下降法更新 W 和 b , 使得输出更好地表示输入。

(2) 展开。在经过多层 RBM 网络预训练后, 编码器和解码器使用得到的权值作为自编码网络的初始权值。将预训练得到的 RBM 网络连接起来并按照自编码网络结构展开。

(3) 权值微调。按照重构误差最小化原则对自编码网络进行调整, 依次经过解码器和编码器利用反向传播算法^[16]对整个自编码网络进行权值微调。通过前向传播的方法, 对每一层的神经元进行前向传导计算, 利用前向传播公式得到各层的激活值。

2.2 自编码网络训练过程

训练阶段主要过程包括以下几个步骤:

(1) 输入可见层变量 $H = \{h_1, h_2, \dots, h_n\}$ 和隐藏层变量 $V = \{v_1, v_2, \dots, v_n\}$, 输出模型参数 $\theta = \{W, \mu, b\}$;

(2) 初始化 RBM 网络权值 $W_{ij} = a_i = b_j = 0$ ($i, j \in Z^+$) 和迭代次数 k ;

(3) 将每个输入特征变量 v_i 赋值给 v^0 ;

(4) 对所有可见单元和隐藏单元, 根据式 7、式 8 求得初始状态和更新状态下的联合概率分布梯度, 并代入式 11 来更新 $\theta = \{W, \mu, b\}$, 即:

$$W_{ij}^{t+1} = W_{ij}^t + p(h_j | v^0) \cdot v_i^0 - p(h_j | v^k) \cdot v_i^k \quad (12)$$

$$a_i^{t+1} = a_i^t + v_i^0 - v_i^k \quad (13)$$

$$b_i^{t+1} = b_i^t + p(h_j | v^0) - p(h_j | v^k) \quad (14)$$

如果 $k = t$, 保存模型参数, 算法结束; 如果 $k > t$, 则 $t = t + 1$, 转向步骤 2。

2.3 权值微调

前后向传播算法 (BP) 是一种有监督算法, 在自编码网络中, BP 算法起到权值微调的作用。首先建立自编码网络的前向传播, 获得模型参数。然后利用反向传播算法按照重构数据误差修改权值, 自顶向下传播至每一层 RBM, 微调整个自编码网络模型参数。输入通过预训练 RBM 得到的参数 θ , 训练调整后的模型参数。初始化 RBM 训练迭代次数 k , 对于每个实例 v_i 输入自编码网络, 计算 v_i 的重构表示 \hat{v}_i , 使得误差沿着自编码网络方向传播。计算每个输出单元 k 的误差项 δ_k :

$$\delta_k = \hat{v}_k' (1 - \hat{v}_k') (v_k - \hat{v}_k) \quad (15)$$

最后, 更新每个网络模型参数 $\theta_{ij} = \theta_{ij} + \Delta \theta_{ij}$ 。如果 $k = t$, 则保存微调后的参数, 算法结束; 如果 $k > t$, $t = t + 1$, 回到步骤 2。

3 GA-FCM 分类器构建

3.1 模糊 C-均值聚类算法 (FCM)

假设 $X = \{x_1, x_2, \dots, x_n\}$ 为 n 维数据样本, c ($2 \leq c \leq n$) 为类别数, $\{L_1, L_2, \dots, L_c\}$ 为分类类别, $U = \{U_1, U_2, \dots, U_n\}$ 为 n 个样本隶属度矩阵。各类别的聚类中心 $v = \{v_1, v_2, \dots, v_c\}$, $\mu_i(x_i)$ 为第 i 个样本对于类别 L_k 的隶属度 (简称为 μ_{ik})。则目标函数 J_b 的表达式为:

$$J_b(U, v) = \sum_{i=1}^n \sum_{k=1}^c (\mu_{ik})^b (d_{ik})^2 \quad (16)$$

其中, $d_{ik} = d(x_i - v_k) = \sqrt{\sum_{j=1}^m (x_{ij} - v_{kj})^2}$, 表示欧几里得距离, 用来度量第 i 个样本 x_i 与第 k 类中心点的距离, m 为样本维度; b 为加权参数, 表示目标函数迭代轻缓程度, 一般为 $2^{[17]}$ 。

计算隶属度 μ_{ik} 和 c 个聚类中心 $\{v_i\}$, 分别对 J_b 求极值有:

$$\mu_{ik} = \frac{1}{\sum_{j=1}^c \left(\frac{d_{jk}}{d_{ik}} \right)^{\frac{b}{1-b}}} \quad (17)$$

$$v_{ij} = \frac{\sum_{k=1}^n (\mu_{ik})^b x_{kj}}{\sum_{k=1}^n (\mu_{ik})^b} \quad (18)$$

其中每个样本对于各个聚类的隶属度和为 1。使用式 17、式 18 反复修改聚类中心和隶属度进行迭代, 使目标函数 J_b 趋于稳定, 完成各类聚类中心和各样本对每个类的隶属度。

3.2 遗传算法实现

遗传算法是一种全局寻优方式, 具有强大的空间搜索能力^[18], 两者结合使用避免 FCM 算法陷入局部

最优。过程如下:

(1) 编码方式: 初始化群体参数, 每个参数对应一个聚类中心。对每个样例进行二进制编码, 对 m 维样本向量, c 个聚类中心的染色体, 待优化的变量数为 $c \times m$ 。假定每个变量使用 k 位二进制编码, 则染色体长度为 $c \times m \times k$ 的二进制编码串。

(2) 适应度函数: 按优胜劣汰的机制, 适应度函数值越小, 繁衍几率越大。排序得到分配函数: $F_{intV} = \text{ranking}(J_k)$

(3) 选择算子: 采用随机遍历抽样(SUS)。

(4) 交叉算子: 采用单点交叉算子。

(5) 变异算子: 以初始概率产生变异基因数, 用随机方法选出变异基因。如果是变异基因, 那么它将进行 0, 1 互变。

3.3 GA-FCM 算法流程

基于遗传算法优化的模糊 C-均值聚类算法模型描述如下:

(1) 初始化 GA-FCM 算法模型参数, 包括模糊 C-均值算法最大迭代次数 N , 目标函数的终止容限 D , 种群个体大小 sizepop , 最大进化次数 MAXGEN , 变异概率 P_m 。

(2) 随机初始化 c 个聚类中心, 并生成初始种群 Chrom, 对每个聚类中心用式 17 计算各样本隶属度, 以及每个个体适应度 f_i , $i = 1, 2, \dots, \text{sizepop}$ 。

(3) 设初始遗传代数循环计数变量 $\text{gen} = 0$ 。

(4) 对群体 Chrom 进行选择、交叉和变异等操作, 对新产生的个体用式 17、式 18 计算 c 个聚类中心和各样本的隶属度, 以及每一个个体的适应度值 f_i 。若 $f_i > f_i$, 则以新个体代替旧个体; 否则接受新个体, 舍弃旧个体。

(5) 若 $\text{Gen} < \text{MAXGEN}$, $\text{Gen} = \text{Gen} + 1$, 转至步骤 4, 否则结束算法。

4 实验结果及分析

4.1 数据预处理

对 KDD'99 数据集^[19]进行预处理作为自编码网络输入数据, 主要包括数值化和归一化过程。

(1) 数值化和归一化。原始 KDD'99 数据集中包含了 41 个特征属性, 其中 3 个为字符型特征属性, 分别为 protocol_type(网络协议类型)、service(目标主机网络服务类型)、flag(连接正常或错误的状态)。protocol_type 包含 3 种协议类型, service 包含 70 种服务类型, flag 包含 11 种状态。文中分别对这 3 种字符型数值进行二进制数值化编码处理, 最后将数据集进行区间 $[0, 1]$ 归一化操作, 得到处理后的数据集。

具体过程见表 1。

表 1 数据集编码

字符类型	字符编码	字符特征属性
TCP	100	protocol_type
UDP	010	protocol_type
ICM	001	protocol_type
aol	1000000...000	service
auth	0100000...000	service
...
Z39_50	0000000...001	service
OTH	10000000000	flag
REJ	01000000000	flag
...
SH	00000000001	flag

(2) 数据集选择。每个数据集随机抽取, 包含正常连接数据样本和入侵数据样本, 如表 2 所示。

表 2 数据集组成

数据集	训练集		测试集	
	正常	异常	正常	异常
D_1	10 004	3 200	9 800	2 300
D_2	10 200	4 500	10 500	4 650
D_3	8 000	3 750	8 500	3 450
D_4	7 800	3 175	7 500	2 250

4.2 模型参数选择实验分析

自编码网络的深度对特征降维有着重要影响, 直接影响数据集维度和分类效果。Hinton 的研究指出 3 层 RBM 网络已能提取有效的特征用于分类任务^[20]。

实验设置 5 种不同 AN-GAFCM 网络结构进行性能对比分析。设置 $\text{AN}^2\text{-GAFCM}$ 、 $\text{AN}^3\text{-GAFCM}$ 、 $\text{AN}^4\text{-GAFCM}$ 、 $\text{AN}^5\text{-GAFCM}$ 、 $\text{AN}^6\text{-GAFCM}$ 的 RBM 网络结构为 122-60-5、122-80-40-5、122-110-70-40-5、122-110-80-50-25-5 和 122-110-80-60-40-20-5。结果表明使用 5 层自编码网络结构性能最佳。

4.3 不同分类器检测性能对比

将 $\text{AN}^5\text{-GA-FCM}$ 模型与传统入侵检测模型做实验对比, 通过常用 PCA 降维算法、SVM 分类算法, 以及神经网络中的 Softmax 分类算法进行实验, $\text{AC}(\%)$ 表示准确率, $\text{TD}(\text{s})$ 表示检测时间, 比较如表 3、图 1 和图 2 所示。

表 3 不同分类器的检测性能对比

算法模型		D_1	D_2	D_3	D_4
$\text{AN}^5\text{-GA-FCM}$	AC	96.89	97.53	97.71	97.35
	TD	2.37	3.03	2.14	2.09
$\text{AN}^5\text{-Softmax}$	AC	90.45	91.43	89.76	89.34
	TD	4.31	4.53	4.29	4.01
$\text{AN}^5\text{-SVM}$	AC	94.78	96.45	95.89	94.56
	TD	3.09	4.38	5.21	4.12
PCA-GAFCM	AC	93.69	92.23	93.47	94.82
	TD	10.38	9.30	10.04	11.39
PCA-SVM	AC	95.69	94.78	95.60	94.59
	TD	8.73	9.48	9.39	8.49

续表 3

算法模型		D_1	D_2	D_3	D_4
FCM	AC	84.74	86.93	85.89	83.59
	TD	3.43	3.40	4.29	3.27
GAFCM	AC	87.39	88.39	85.28	84.89
	TD	2.39	2.38	1.98	2.08

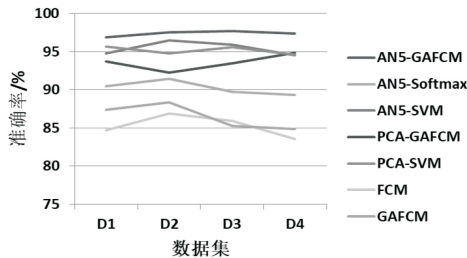


图1 不同算法的正确率

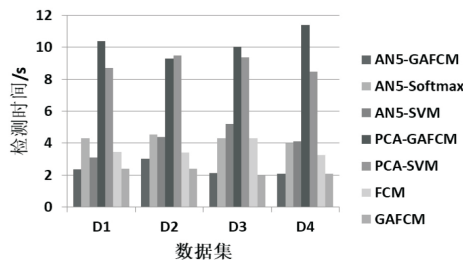


图2 不同算法的检测时间

结果表明,文中提出的自编码网络结合遗传算法优化的模糊C-均值算法模型具有较高的检测率和较低的检测时间。应用自编码网络降维效果优于传统PCA降维算法,使用GA-FCM算法分类在检测时间和检测率上优于Softmax、SVM、FCM。

5 结束语

检测入侵异常行为,需要检测庞大复杂的数据流量,特征选择的好坏直接影响检测效果。针对网络数据流量的稀疏多样性等特点,将自编码网络特征降维深度学习用于处理入侵检测数据方面,并提出一种改进缩短检测时间的优化模糊聚类算法GA-FCM。通过分类算法和降维算法的对比,证明AN-GA-FCM算法模型具有较高的检测率和较短的运行时间。

参考文献:

- [1] DENATIOUS D K, JOHN A. Survey on data mining techniques to enhance intrusion detection[C]//International conference on computer communication and informatics. Coimbatore, India: IEEE, 2012: 1-5.
- [2] CHITRAKAR R, HUANG Chuanhe. Selection of candidate support vectors in incremental SVM for network intrusion detection[J]. Computers and Security, 2014, 45: 231-241.
- [3] SRINIOY S, KURUTACH W, CHIMPHLEE W, et al. Intrusion detection via independent component analysis based on rough fuzzy[J]. WSEAS Transactions on Computers, 2006, 5

- (1): 43-48.
- [4] HATHAWAY R J, BEZDEK J C. Local convergence of the fuzzy c-means algorithms[J]. Pattern Recognition, 1986, 19(6): 477-480.
- [5] WANG Qiang, MEGALOOIKONOMOU V. A clustering algorithm for intrusion detection[J]. Proceedings of SPIE, 2008, 5812: 31-38.
- [6] OH S H, LEE W S. An anomaly intrusion detection method by clustering normal user behavior[J]. Computers & Security, 2003, 22(7): 596-612.
- [7] 黄思慧, 陈万忠, 李 晶. 基于PCA和ELM的网络入侵检测技术[J]. 吉林大学学报: 信息科学版, 2017, 35(5): 576-583.
- [8] KUANG Fangjun, XU Weihong, ZHANG Siyang. A novel hybrid KPCA and SVM with GA model for intrusion detection[J]. Applied Soft Computing, 2014, 18: 178-184.
- [9] HOZ E D L, HOZ E D L, ORTIZ A, et al. PCA filtering and probabilistic SOM for network intrusion detection[J]. Neurocomputing, 2015, 164: 71-81.
- [10] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504-507.
- [11] ZABALZA J, REN Jinchang, ZHENG Jiangbin, et al. Novel segmented stacked autoencoder for effective dimensionality reduction and feature extraction in hyperspectral imaging[J]. Neurocomputing, 2016, 185: 1-10.
- [12] 刘 忠, 茆诗松. 分组数据的Bayes分析—Gibbs抽样方法[J]. 应用概率统计, 1997, 13(2): 211-216.
- [13] BENGIO Y, DELALLEAU O. Justifying and generalizing contrastive divergence[M]. [s.l.]: MIT Press, 2009.
- [14] 高 妮. 网络安全多维动态风险评估关键技术研究[D]. 西安: 西北大学, 2016.
- [15] 胡 洋. 基于马尔可夫链蒙特卡罗方法的RBM学习算法改进[D]. 上海: 上海交通大学, 2012.
- [16] LIU Fan, XU Feng, YANG Sai. A flood forecasting model based on deep learning algorithm via integrating stacked autoencoders with BP neural network[C]//IEEE third international conference on multimedia big data. Laguna Hills, CA, USA: IEEE, 2017: 58-61.
- [17] 高新波, 裴继红, 谢维信. 模糊c-均值聚类算法中加权指数m的研究[J]. 电子学报, 2000, 28(4): 80-83.
- [18] 吉根林. 遗传算法研究综述[J]. 计算机应用与软件, 2004, 21(2): 69-73.
- [19] TAVALLAEI M, BAGHERI E, LU Wei, et al. A detailed analysis of the KDD CUP 99 data set[C]//IEEE international conference on computational intelligence for security & defense applications. Ottawa, ON, Canada: IEEE, 2009: 1-6.
- [20] HINTON G E, OSINDERO S, TEH Y W. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7): 1527-1554.