基于差分 WGAN 的网络安全态势预测

王婷婷 朱 江
(重庆邮电大学通信与信息工程学院 重庆 400065)

摘 要 文中提出了一种基于差分 WGAN(Wasserstein-GAN)的网络安全态势预测机制,该机制利用生成对抗网络(Generative Adversarial Network,GAN)来模拟态势的发展过程,从时间维度实现态势预测。为了解决 GAN 具有的网络难以训练、collapse mode 及梯度不稳定的问题,提出了利用 Wasserstein 距离作为 GAN 的损失函数,并采用在损失函数中添加差分项的方法来提高态势值的分类精度,同时还证明了差分 WGAN 网络的稳定度。实验结果与分析表明,该机制相比其他机制而言,在收敛性、预测精度和复杂度方面具有优势。

关键词 态势感知,态势预测,生成对抗网络,差分,Wasserstein-GAN
中图法分类号 TN918.1 **文献标识码** A

Network Security Situation Forecast Based on Differential WGAN

WGAN Ting-ting ZHU Jiang
(School of Communication and Information Engineering,Chongqing University of Post and Telecommunications,Chongqing 400065,China)

Abstract A network security posture prediction mechanism based on differential WGAN(Wasserstein-GAN) is presented in this paper. This mechanism uses Generative adversarial network (GAN) to simulate the development process of the situation,and realizes the situation forecast from the time Dimension. In order to solve the problem of difficult network training,collapse mode and gradient instability of GAN,this paper put forward the method by using Wasserstein distance as the loss function of GAN and adding the difference term in the loss function,to improve the classification precision of the situation value. The stability of the differential WGAN network was also proved. Experimental and analysis results show that this mechanism has advantages over other mechanisms in terms of convergence,accuracy and complexity.

Keywords Situational awareness,Situation forecast,Generative adversarial network,Difference,Wasserstein-GAN

1 引言

现今,全球的网络空间遭遇了巨大的安全挑战,国家型黑客攻击事件频发、针对关键基础设施与物联网的攻击不断、勒索软件盛行、数据泄露严重等。未来的网络安全将会如何发展,如何能准确预测网络安全态势是未来的研究重点。网络安全态势预测是网络安全态势感知(Network Security Situation Awareness,NSSA)[1]的最终目的。

在网络安全领域,态势预测已成为热点。网络安全态势预测就是根据一段时间内的网络安全数据即态势评估得到的态势值,运用专家知识及数据挖掘等理论方法分析预测未来时间的网络态势可能的发展趋势,使安全管理员能够在可能的安全攻击前做好准备。随着机器学习算法的不断发展,网络安全态势预测主要是基于 D-S 证据理论[2]、支持向量机(Support Vector Machine,SVM)[3]、深度信念网络(Deep Belief Network,DBN)[4]等理论的预测方法。虽然这些方法也取得了一定的效果,但仍需要不断地完善。随着机器学习的不断进步和人工智能的不断发展,许多专家尝试着在新的领域预测网络安全态势,机器学习在近几年是大家研究的重点对象,在图像分类、可视化等方面效果卓著,在网络安全方面还需要不断地探索和创新。

随着机器学习研究的不断深入,生成对抗网络[5]是基于可微生成网络的另一种生成式,训练 GAN 需要达到纳什均衡[6],训练 GAN 模型是不稳定的。在 GAN 模型的基础上也做了很多改进,如 DCGAN[7]依靠的是对判别器和生成器的架构进行实验枚举,最终找到一组较好的网络架构设置,但实际上这种方法没有彻底解决问题,而 Wasserstein-GAN(WGAN)[8]得到了很好的效果。本文将 WGAN 运用到网络安全中,并在损失函数中添加差分项,提出了一种基于差分 WGAN 的态势预测方法。该方法充分考虑了不同态势要素的依赖关系,利用态势要素时间维度上的关联性对未来的网络安全态势要素进行预测,更客观地反映了历史网络安全态势对未来态势的影响。

2 基于差分 WGAN 的态势预测方法

2.1 差分 WGAN 流程图

生成式可以通过真实数据的本质特征来刻画样本的数据分布特征,生成与训练样本相似的新数据。GAN 是由 Goodfellow 等于 2014 年提出的一种生成模型,不同于传统的生成模型,其在网络结构上除了生成网络,还包含一个判别网络。

王婷婷(1993-),女,硕士生,主要研究方向为网络安全态势感知,E-mail:1879213049@163.com;朱 江(1977-),男,博士,教授,主要研究方向为认知无线电。

生成网络与判别网络之间是一种对抗的关系，对抗源自于博弈论[9]的思想，博弈双方在平等的对局中各自利用对方的策略来变换自己的对抗策略，以此达到某个最优的状态。将这种博弈的方式引用到生成对抗网络中的生成器和判别器中，将其视为博弈双方。生成器拟合数据的产生过程生成模型样本，使生成器估测到数据样本的分布。

差分 WGAN 由两个模型构成[10]，即生成模型 G 和判别模型 D，随机噪声 Z 通过 G 生成尽量服从真实数据分布 p_{data} 的样本 $G(z)$，判别模型 D 通过损失函数添加了差分项的损失函数，可以判断出输入样本是真实数据 x 还是生成数据 $G(z)$。G 和 D 都可以是非线性的映射函数，如多层感知器。差分 WGAN 的流程如图 1 所示。

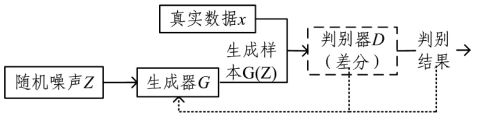

图 1 差分 WGAN 流程图

2.2 WGAN 核心原理描述

在生成器给定的情况下需要优化判别器。训练判别器是实现最小化交叉熵的过程。$E(\cdot)$ 为期望值的计算，x 采样于真实数据分布 $p_{\text{data}}(x)$，z 采样于先验分布 $p_z(z)$。生成器为了学习数据 x 的分布，由先验噪声分布 $p_z(z)$ 构建了一个映射空间 $g(Z;\theta G)$，所对应的判别器映射函数为 $D(x;\theta d)$，输出一个标量表示 x 为真实数据的概率为：

$$\min_G \max_D V(D,G)=E_{x\sim P_{\text{data}(x)}}[\log D(x)]+E_{z\sim P_{z(z)}}[\log(1-D(G(z)))] \tag{1}$$

将式(1)拆分为两个部分来理解：1)$E_{x\sim P_{\text{data}(x)}}[\log D(x)]$，其中，$x$ 表示真实样本，$D(x)$ 表示 x 通过判别网络判断其为真实样本的概率；2)$E_{z\sim P_{z(z)}}[\log(1-D(G(z)))]$，$z$ 表示输入生成样本的噪声，$G(z)$ 表示生成网络由噪声 z 生成的样本，而 $D(G(z))$ 表示生成样本通过判别网络后，判断其为真实样本的概率。生成网络要求生成样本越接近真实则样本越好，即 $D(G(z))$ 越接近 1 越好，这时 $V(D,G)$ 会变小；而判别网络的目的是让 $D(x)$ 接近 1，而 $D(G(z))$ 接近 0，此时 $V(D,G)$ 会增大。

首先由式(1)可以得到，在生成器 G 固定参数时计算判别器 D 的最优状态。对于一个具体的样本 x，它可能来自真实分布也可能来自生成分布，它对式(1)损失函数的贡献是：

$$-P_r(x)[\log D(x)]-P_g(x)[\log(1-D(x)] \tag{2}$$

令其关于 $D(x)$ 的导数为 0，得：

$$-\frac{P_r(x)}{D(x)}+\frac{P_g(x)}{1-D(x)}=0 \tag{3}$$

化简得最优判别器为：

$$D^*(x)=\frac{P_r(x)}{P_r(x)+P_g(x)} \tag{4}$$

这个结果从直观上很容易理解，就是参考一个样本 x 来自真实分布和生成分布的可能性的相对比例。如果 $P_r(x)=0$ 且 $P_g(x)\neq 0$，最优判别器就应该非常自信地给出概率 0；如果 $P_r(x)=P_g(x)$，说明该样本是真是假的可能性刚好一半，此时最优判别器也应该给出概率 0.5。

代入式(1)，再进行简单的变换可以得到：

$$\mathrm{E}_{x\sim P_{\text{data}(x)}}\log\frac{P_r(x)}{\frac{1}{2}[P_r(x)+P_g(x)]}+\mathrm{E}_{z\sim P_{z(z)}}\log\frac{P_g(x)}{\frac{1}{2}[P_r(x)+P_g(x)]}-2\log 2 \tag{5}$$

变换成这个样子是为了引入 Kullback Leibler divergence（KL 散度）[11]和 Jensen Shannon divergence(JS 散度)[12]这两项重要的相似度衡量指标：

$$KL(P_1\parallel P_2)=E_{x\sim P_1}\log\frac{P_1(x)}{P_2} \tag{6}$$

$$JS(P_1\parallel P_2)=\frac{1}{2}KL(P_1\parallel\frac{P_1+P_2}{2})+\frac{1}{2}KL(P_2\parallel\frac{P_1+P_2}{2}) \tag{7}$$

于是式(5)可以写成：

$$2JS(P_1\parallel P_2)-2\log 2 \tag{8}$$

在(近似)最优判别器下，最小化生成器的 loss 等价于最小化 $P_r(x)$ 与 $P_g(x)$ 之间的 JS 散度，而由于 $P_r(x)$ 与 $P_g(x)$ 几乎不可能有不可忽略的重叠，因此无论它们相距多远，JS 散度都是常数 log2，最终导致生成器的梯度(近似)为 0，梯度消失。在 KL 散度条件下存在梯度不平衡以及惩罚不平衡导致 mode collapse[13]等问题。

针对 GAN 的梯度消失、梯度不平衡以及惩罚不平衡导致的 mode collapse 等问题，在 GAN 中引入了差分 Wasserstein 距离[14]作为损失函数，由于它相对 KL 散度与 JS 散度具有优越的平滑特性，理论上可以解决梯度消失问题。接着通过数学变换将 Wasserstein 距离写成可求解的形式，利用一个参数数值范围受限的判别器神经网络来最大化这个形式，就可以近似 Wasserstein 距离，从而有效拉近生成分布与真实分布。

2.3 差分 WGAN 算法描述

WGAN 最大的贡献在于利用 Wasserste-in 距离代替 GAN 中的 JS 散度或 KL 散度，极大地缓解了 GAN 难以训练的问题，Wasserstein 距离又为 Earth-Mover(EM)距离，定义如下：

$$W(P_r,P_g)=\inf_{\gamma\sim\prod(P_r,P_g)}\mathrm{E}_{(x,y\sim\gamma)}[\parallel x-y\parallel] \tag{9}$$

其中，$\prod(P_r,P_g)$ 为 P_r 和 P_g 组合起来的所有可能的联合分布的集合，反过来说 $\prod(P_r,P_g)$ 中的每一个分布的边缘分布都是 P_r 和 P_g。对于每一个可能的联合分布 γ 而言，可以从中采样 $(x,y)\sim\gamma$ 得到一个真实样本 x 和一个生成样本 y，并计算出这对样本的距离 $\parallel x-y\parallel$，因此可以计算该联合分布 γ 下样本对距离的期望值 $\mathrm{E}_{(x,y\sim\gamma)}[\parallel x-y\parallel]$。在所有可能的联合分布中将能够对这个期望值取到的下界 $\inf_{\gamma\sim\prod(P_r,P_g)}\mathrm{E}_{(x,y\sim\gamma)}[\parallel x-y\parallel]$ 定义为 Wasserst-ein 距离。

因为 Wasserstein 距离定义式(9)中的 $\inf_{\gamma\sim\prod(P_r,P_g)}$ 无法直接求解，可以用一个已有的定理把它变换为：

$$W(P_r,P_g)=\frac{1}{K}\sup_{\parallel f\parallel_L\leqslant K}\mathrm{E}_{x\sim P_r}[f(x)]-\mathrm{E}_{x\sim P_g}[f(x)] \tag{10}$$

这个过程已被文献[15]证明。首先需要介绍一个概念——Lipschitz 连续[16]。它其实就是在一个连续函数 f 中额外施加了一个限制，要求存在一个常数 $K\geqslant 0$ 使得定义域内的任意两个元素 x_1 和 x_2 都满足：

$$|f(x_1)-f(x_2)|\leqslant K|x_1-x_2| \tag{11}$$

此时称函数 f 的 Lipschitz 常数为 K。

简单来说，若 f 的定义域是实数集合，则上面的要求就等价于 f 的导函数绝对值不超过 K，又如 $\log(x)$ 不是 Lipschitz 连续，因为它的导函数没有上界。Lipschitz 连续条件限制了一个连续函数的最大局部变动幅度。特别地，可以用一组参数 w 来定义一系列可能的函数 f_w，此时求解式(9)可以近似变成求解如下形式：

$$KW(P_r,P_g)=\max_{w:\|f_w\|_L\leqslant K}\mathrm{E}_{x\sim P_r}[f_w(x)]-\mathrm{E}_{x\sim P_g}[f_w(x)] \tag{12}$$

本文构造了一个含参数 w、最后一层不是非线性激活层的判别器网络 f_w，在限制 w 不超过某个范围的条件下使得判别器的损失函数为：

$$L=-\mathrm{E}_{x\sim P_r}[f_w(x)]+\mathrm{E}_{x\sim P_g}[f_w(x)] \tag{13}$$

Lipschitz 限制要求判别器的梯度不超过 $K(K=1)$，可以在公式末尾添加一个差分项，损失函数为：

$$L=-\mathrm{E}_{x\sim P_r}[D_w(x)]+\mathrm{E}_{x\sim P_g}[D_w(x)]+\lambda\mathrm{E}_{x_1\sim P_{\hat{x}},x_2\sim P_{\hat{x}}}\left[\frac{|D(x_1)-D(x_2)|}{\|x_1-x_2\|}-1\right]^2 \tag{14}$$

也就是说，我们仍然是在分布 $p_{\hat{x}_1}$ 上随机采样，但是一次采两个，然后要求它们的连线斜率接近 1，以此来限制真假样本之间的距离。

定理 1　在 WGAN 损失函数中，使用 Lipschitz 连续，在末尾添加差分方程作为新的损失函数，在分布 $p_{\hat{x}_1}$ 上随机采样，存在一个常数 $K=1$ 使得定义域内的任意两个元素 x_1 和 x_2 都满足式(11)，可以有效地稳定梯度的变化。

判别器都想尽可能地拉大真实样本与生成样本之间的距离，不加差分限制项，通常添加 weight clipping 进行限制，但其是对样本空间全局生效，难免会导致梯度消失或者梯度爆炸；而差分只针对真假样本集中区域，而且把梯度限制在 1 附近，可控性非常强。

证明：已知 $x_r\sim p_{x_r}$，$x_g\sim p_{x_g}$，并假设 $\varepsilon\sim Uniform[0,1]$ 在 x_r，x_g 中间随机插值，即：

$$\hat{x}=\varepsilon x_r+(1-\varepsilon)x_g \tag{15}$$

此时 $\hat{x}$ 满足的分布记为 $p_{\hat{x}_r}$，在 $p_{\hat{x}_r}$ 上随机取样，在其中选取两个不一样的值，例如 $x_1\sim p_{\hat{x}}$ 和 $x_2\sim p_{\hat{x}}$，这两个值都是真实样本和生成样本的集中区域选取的，控制它们之间的距离，对其进行限制，可以防止距离过大或者距离过小，为判别器分辨带来了很好的效果。将 weight clipping 和差分做简单的对比，明显看到了差分的优势。

由图 2 可知，用了差分项之后，梯度值变化甚微，这会给判别器对真实样本和生成样本的区分带来意想不到的效果。

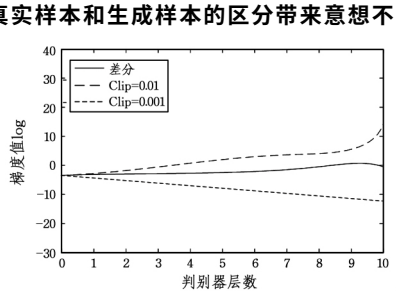

图 2　梯度值比较

差分 WGAN 成功地解决了 GAN 遇到的问题：解决了 GAN 训练不稳定的问题，不再需要小心平衡生成器和判别器的训练程度，同时基本解决了 collapse mode 的问题，确保了生成样本的多样性。

WGAN 的问题在于，判别器是一个多层网络。在处理 Lipschitz 限制条件时直接采用了 weight clipping，然而这种方法会让参数限制在 clip 的范围内，因此针对性地给出了改进要点，在损失函数后面加了一个可以进行限制的差分项，来限制参数的范围。具体算法如算法 1 所示。

算法 1　差分 WGAN 算法

输出：$\lambda=10$，$\alpha=0.0001$，$\beta_1=0$，$\beta_2=0.9$，$n=5$，$m=64$

输出：w 和 θ

While θ do

for t=0,…,n do

样本 $\{x^{(i)}\}_{i=1}^m\sim P_r$ 为真实数据

样本 $\{z^{(i)}\}_{i=1}^m\sim p(z)$ 为生成样本

$x_r\sim p_{x_r}$，$x_g\sim p_{x_g}$

$\varepsilon\sim$ Uniform[0,1]

$\hat{x}=\varepsilon x_r+(1-\varepsilon)x_g$

$L^i=-\mathrm{E}_{x\sim P_r}[D_w(x)]+\mathrm{E}_{x\sim P_g}[D_w(x)]+\lambda\mathrm{E}_{x_1\sim P_{\hat{x}},x_2\sim P_{\hat{x}}}[\frac{|D_w(x_1)-D_w(x_2)|}{\|x_1-x_2\|}-1]^2$

end for

$w\leftarrow \mathrm{Adam}(\nabla w\frac{1}{m}\sum_{i=1}^m L^i,w,\alpha,\beta_1,\beta_2)$

$\theta\leftarrow \mathrm{Adam}(\nabla\theta\frac{1}{m}\sum_{i=1}^m -D_w(G_\theta(Z)),w,\alpha,\beta_1,\beta_2)$

end while

算法 1 中，λ 是惩罚系数，α 是学习率，n 为迭代次数。m 是批量值，Adam 算法的高阶参数 β_1，β_2，w_0 为权重初始值，θ_0 为生成器初始值。

3　态势预测流程图

差分 WGAN 预测模型中，生成器 G 和判别器 D 本文中用的卷积神经网络(Convolution neural network，CNN)[19] 也是一种深度卷积对抗生成网络，具体模型如图 3 所示。

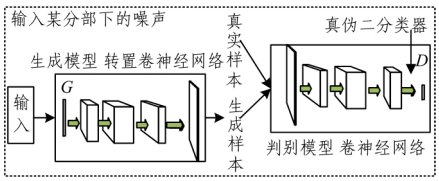

图 3　深度卷积对抗生成网络

差分 WGAN 的网络预测模型中的 CNN 有一些改变，分别为：

(1)取消所有 pooling 层。G 网络中使用转置卷积进行上采样，D 网络中采用加入步长的卷积代替 pooling。

(2)去掉 FC 层，使网络变为全卷积网络。

(3)G 网络中使用 ReLU 作为激活函数，最后一层使用 tanh。

(4)D 网络中，使用 LeakyReLU 作为激活函数。

实验中，将前 xx 天的数据作为生成器的输入，生成一种分布，后 xx 天的数据作为真实数据输入到判别器中，判别器会判别前 xx 天的数据分布与后 xx 天的数据进行区分，不断地更新网络来达到概率近似为 1 的状态，也就是预测前 xx 天的态势是不是会发展成后××天的态势，这样也就达到了预测的效果，可以知道未来××天内的发展态势。差分 WGAN 安全态势预测流程如图 4 所示。

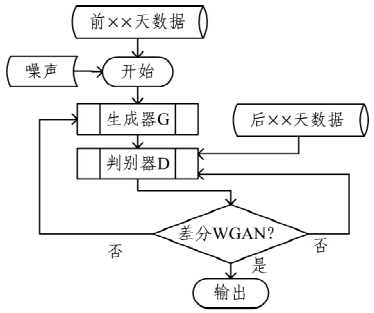

图 4 WGAN 预测流程图

4 实验结果及分析

4.1 实验数据及样本选取

分析安全数据的攻击特点，其具有不确定性和连续性，本文选用某公司 2013 年 7 月到 9 月中的 95 天的防火墙、IDS 等历史日志信息作为原始数据集，对每天的日志信息进行一次采样。通过文献[21]的网络安全态势评估方法从服务、主机、网络系统 3 个层次对攻击威胁和服务节点权重进行量化计算，从而得到网络安全态势值，其简化后的量化模型如图 5 所示。

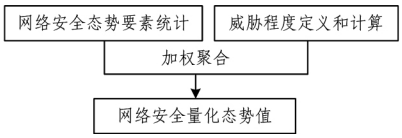

图 5 网络安全态势评估量化

由于安全态势值是随机的，量纲差异大，为了提升模型的训练速度，对态势值进行极值标准化处理，处理公式如下：

$$\hat{X}=\frac{X-X_{\min}}{X_{\max}-X_{\min}} \tag{16}$$

其中，$X_{\min}$和 $X_{\max}$为样本中最小和最大的态势值。X 和$\hat{X}$分别为处理前后的态势值。极值标准化后的网络安全态势数据如图 6 所示。

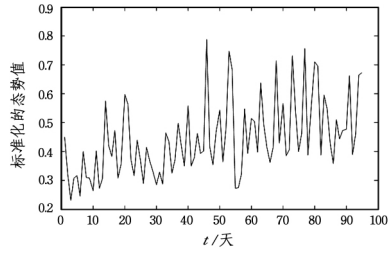

图 6 极值化后的网络安全态势值

将态势评估值得到的一维时间序列样本拓阶处理，确定拓阶维数为 5，再对训练样本一维时间序列运用滑动窗口进行数据重构。重构结果如表 1 所列。

表 1 训练数据重构结果

输入样本	输出样本
X_1, X_2, X_3, X_4, X_5	X_6
X_2, X_3, X_4, X_5, X_6	X_7
…	…
$X_{74}, X_{75}, X_{76}, X_{77}, X_{78}$	X_{79}
…	…

4.2 实验结果分析与对比

4.2.1 收敛分析

将数据重构后规范在 tanh 的[−1,1]。mini-batch 训练中 batch size 为 128。所有的参数初始化由(0,0.02)的正态分布中随机得到 LeakyReLU 的斜率是 0.2。

将加了差分项的损失函数与不加差分项的损失函数进行比较，由图 7 可知，加了差分项的误差更小。

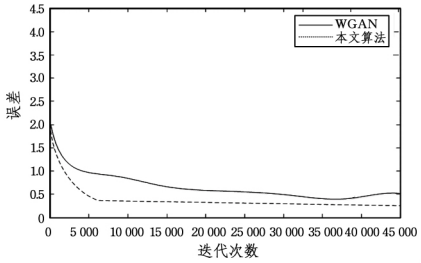

图 7 误差随迭代次数的变化曲线

4.2.2 与其他预测方法的对比

将差分 WGAN 的预测方法和常见的 GAN 改进方法进行比较，如 WGAN，DCGAN 以及 LSGAN 等，结果如图 8 所示。

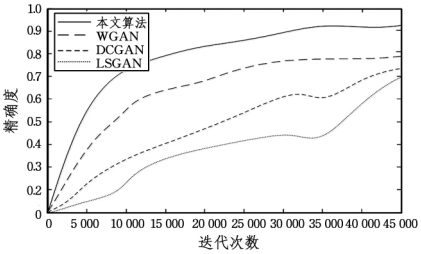

图 8 不同预测方法的态势值对比

由图 8 可以看出，差分 WGAN 的预测方法的效果很好。这是因为加了差分项的损失函数可控性强，且损失函数的 Wasserstein 距离解决了 GAN 中梯度消失、梯度不稳定以及 collapse mode 等问题。

4.2.3 复杂度比较

历史态势要素样本数为 n，batch 长度为 m，每层迭代次数为 $Q=n/m$(为整数)，判别器隐含层次数为 K，反向微调判别网络次数为 M，生成器隐含层数为 N。不同方法的复杂度对比如表 2 所列。可以看出，各方法的量级是相同的，只是具体的反向微调的次数 M 不一样，并没有成规模的增加，但实验仍取得了很好的效果。

表 2 不同方法复杂度对比

方法	复杂度
本文方法	$O((M_1+1)(Q\times N\times K))$
WGAN	$O((M_2+1)(Q\times N\times K))$
DCGAN	$O((M_3+1)(Q\times N\times K))$
LSGAN	$O((M_4+1)(Q\times N\times K))$

结束语 本文提出了一种基于差分 WGAN 的网络安全态势预测方法，根据实际网络环境建立循环神经网络模型，提取网络安全态势要素训练模型并对未来网络安全变化趋势进行预测。采用某公司 7 月到 9 月中的 95 天的防火墙、IDS 等历史日志信息作为原始数据集进行实验，结果表明了该方法具有可行性以及较高的准确度。

参 考 文 献

[1] 谢丽霞，王亚超，于巾博. 基于神经网络的网络安全态势感知

[J]. 清华大学学报(自然科学版)，2013(12)：1750-1760.
[2] 石波，谢小权. 基于 D-S 证据理论的网络安全态势预测方法研究[J]. 计算机工程与设计，2013，34(3)：821-825.
[3] 陈善学，杨政，朱江，等. 一种基于累加 PSO-SVM 的网络安全态势预测模型[J]. 计算机应用研究，2015，32(6)：1778-1781.
[4] 田庆安，郭玉锦，王文涛. 基于小波与 DBN 的负荷预测模型[J]. 兰州理工大学学报，2017，43(2)：110-114.
[5] ZHANG H，XU T，LI H. StackGAN：Text to Photo-Realistic Image Synthesis with Stacked Generative Adversarial Networks [C]// IEEE International Confernece on Computer Vision(ICCV)，2016：5908-5916.
[6] 吴昊. 随机森林预测与纳什均衡策略的高职英语教学模式研究[J]. 佳木斯职业学院学报，2017(2).
[7] CHANG J，SCHERER S. Learning representations of emotional speech with deep convolutional generative adversarial networks [C]// IEEE International Conference on Acoustics，Speech and Signal Processing. IEEE，2017：2746-2750.
[8] ZHAO Y，TAKAKI S，LUONG H T，et al. Wasserstein GAN Waveform Loss-based acoustic model training form Multi-speaker Text-toSpeechsynthesis systems using a wav Net rocoder[J]. IEEE Access，2017：7(1)：1-10.
[9] 王峰，何俊. 博弈论在通信对抗态势预测中的应用[J]. 运筹与管理，2011，20(2)：132-136.
[10] RADFORD A，METZ L，CHINTALA S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks[J]. Computer Science，2015.
[11] 朱红春，黄伟，刘海英，等. 基于 KL 散度的面向对象遥感变化检测[J]. 国土资源遥感，2017，29(2)：46-52.
[12] 张妍，韩光威，陆宁云，等. 基于 JS 散度的轨道车辆门系统健康状态评估方法[J]. 机械设计与制造工程，2017，46(11)：122-127.
[13] SRIVASTAVA A，VALKOV L，RUSSELL C，et al. VEEGAN：Reducing Mode Collapse in GANs using Implicit Variational Learning[J]. arXiv：1705.07761，2017.
[14] 王群，董文略，杨莉. 基于 Wasserstein 距离和改进 K-medoids 聚类的风电/光伏经典场景集生成算法[J]. 中国电机工程学报，2015，35(11)：2654-2661.
[15] ARJOVSKY M，CHINTALA S，BOTTOU L. Wasserstein GAN [J]. 2017.
[16] NAKAJO K. Improved gradient method for monotone and lipschitz continuous mappings in banach spaces[J]. Acta Mathematica Scientia(English Series)，2017，37(2)：342-354.
[17] 李南星，盛益强，倪宏. 基于 LM 算法的 MLP 模型及其应用[J]. 网络新媒体技术，2018，7(1)：59-63.
[18] MUKKAMALA M C，HEIN M. Variants of RMSProp and Adagrad with Logarithmic Regret Bounds[J]. IEEE Transactions Biomed Engineering，2017，5(6)：1220-1228.
[19] QIU Z，YAN Z，FEI Y，et al. RGB-DI Images and Fall Convolution Nuernal Network-Based Outdoor Scene Understanding for Mobile Robots[J]. IEEE Transactions on Instrumentation & Measurement，2018，1(99)：1-11.
[20] IOFFE S，SZEGEDY C. Batch Normalization：Accelerating Deep Network Training by Reducing Internal Covariate Shift[J]. arXiv：1502.03167，2015.
[21] 陈秀真，郑庆华，管晓宏，等. 层次化网络安全威胁态势量化评估方法[J]. 软件学报，2006，17(4)：885-897.

(上接第 413 页)

[10] KIM H，JEON W，LEE K，et al. Cryptanalysis and improvement of a biometrics-based multi-server authentication with key agreement scheme[C]// Proc. of the 12th Int'l Conf. on Computational Science and Its Applications (ICCSA 2012). IEEE，2012：391-406.
[11] HE D B，WANG D. Robust biometrics-based authentication scheme for multi-serve environmental[J]. IEEE Systems Journal，2005，9(3)：816-823.
[12] ODELU V，DAS A K，GOSWAMI A. Crytanalysis on robust biometrics-based authentication scheme for multi-server environment [EB/OL]. http://eprint.iacr.org/2014/715.
[13] CHUANG M C，CHEN M C. An anonymous multi-server authenticated key agreement scheme based on trust computing suing smart cards and biometric [J]. Expert Systems with Applications，2014，41(4)：1411-1418.
[14] MISHRA D，DAS A，MUKHOPADHYAY S. A secure user anonymity-preserving biometric-based multi-sever authenticated key agreement scheme using smart cards [J]. Expert Systems with Applications，2014，41(18)：8129-8143.
[15] 王瑞兵，陈建华，张媛媛. 一个匿名的基于生物特征的多服务器的密钥认证协议方案的研究[J]. 计算机应用研究，2016，33(7)：2190-2196.
[16] CHAUDHRY S A. A secure biometric based based multi-server authentication scheme for social multimedia network [J]. Multimedia Tools & Applications，2016，75(20)：1-21.
[17] XIA P Z，CHEN J H. Three-factor authentication scheme for multi-servers environments based on elliptic curve cryptography [J]. Application Research of Computers，2017，34(10)：3061-3067.
[18] 殷秋实，陈建华. 多服务器环境下基于椭圆曲线密码的改进的身份认证协议[J]. 计算机科学，2018，45(6)：111-116.
[19] 汪定，李文婷，王平. 对三个多服务器环境下匿名认证协议的分析[J]. 软件学报，2018，29(7)：1937-1952.
[20] 汪定，马春光，翁臣，等. 一种适于受限资源环境的远程用户认证方案的分析与改进[J]. 电子与信息学报，2012，34(10)：2520-2526.
[21] WAN T，LIU Z X，MA J F. Authentication and key agreement protocol for multi-server architecture[J]. Journal of Computer Research and Development，2016，53(11)：2446-2453.
[22] AMIN R. Cryptanalysis and efficient dynamic ID based remote user authentication scheme in multi-server environment ssing smart card[J]. Int'l Journal of Network Security，2016，18(1)：172-181.
[23] REDDY A G，YOON E J，DAS A K，et al. Design of mutually authenticated key agreement protocol resistant to impersonation attacks for multi-server environment[J]. IEEE Access，2017，5：3622-3639.