基于APDERBF神经网络的网络安全态势预测方法

申请号：201610871705.X
申请日：2016-09-30

申请(专利权)人　重庆邮电大学
地址　400065 重庆市南岸区南山街道崇文路2号
发明(设计)人　李方伟　李骐　李俊瑶

主分类号　H04L29/06(2006.01)I
分类号　H04L29/06(2006.01)I　H04L12/24(2006.01)I　G06N3/04(2006.01)I
公开(公告)号　106411896A
公开(公告)日　2017-02-15
专利代理机构　北京一格知识产权代理事务所(普通合伙) 11316
代理人　滑春生

(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 106411896 A
(43)申请公布日 2017.02.15

(21)申请号 201610871705.X

(22)申请日 2016.09.30

(71)申请人 重庆邮电大学
地址 400065 重庆市南岸区南山街道崇文路2号

(72)发明人 李方伟 李骐 李俊瑶

(74)专利代理机构 北京一格知识产权代理事务所(普通合伙) 11316
代理人 滑春生

(51)Int.Cl.
H04L 29/06(2006.01)
H04L 12/24(2006.01)
G06N 3/04(2006.01)

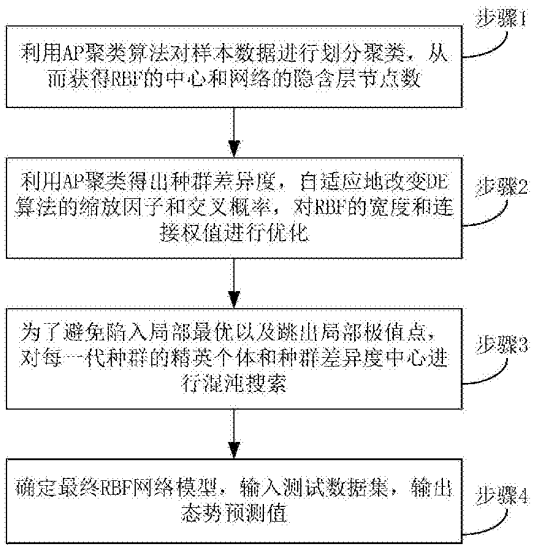
权利要求书3页 说明书8页 附图4页

(54)发明名称

基于APDE-RBF神经网络的网络安全态势预测方法

(57)摘要

本发明属于网络安全技术领域，特别涉及一种基于吸引子传播差分进化算法的径向基函数APDE-RBF神经网络的网络安全态势预测方法，包括利用AP聚类算法对样本数据进行划分聚类，从而获得径向基函数RBF的中心和网络的隐含层节点数；利用AP聚类得出种群差异度，自适应地改变DE算法的缩放因子和交叉概率，对RBF的宽度和连接权值进行优化；同时为了避免陷入局部最优以及跳出局部极值点，对每一代种群的精英个体和种群差异度中心进行混沌搜索；确定最终RBF网络模型，输入测试数据集，输出态势预测值；本发明旨在增强泛化能力的同时，提高对网络安全态势的预测精度。



CN 106411896 A

1.一种基于吸引子传播差分进化算法的径向基函数APDE-RBF神经网络的网络安全态势预测方法，其特征在于，该方法包括以下步骤：

步骤1：利用吸引力传播AP聚类算法对样本数据进行划分聚类，从而获得径向基函数RBF的中心和网络的隐含层节点数；

步骤2：利用AP聚类得出种群差异度，自适应地改变差分进化DE算法的缩放因子和交叉概率，对RBF的宽度和连接权值进行优化；

步骤3：为了避免陷入局部最优以及跳出局部极值点，对每一代种群的精英个体和种群差异度中心进行混沌搜索；

步骤4：确定最终RBF网络模型，输入测试数据集，输出态势预测值。

2.根据权利要求1所述的基于APDE-RBF神经网络的网络安全态势预测方法，其特征在于，所述步骤1进一步包括以下步骤：

步骤11：利用欧氏距离计算输入节点之间的相似度矩阵S为：$S(i,k)=-||x_i-x_k||^2$，其中x_i和x_k表示RBF神经网络任意两个输入节点，S(i,k)表示点x_k作为点x_i的聚类中心的相似度，其值储存于相似矩阵S中；

步骤12：初始化吸引度矩阵R和归属度矩阵A为R(i,k)＝0，A(i,k)＝0，其中R(i,k)表示点x_k适合作为数据点x_i的聚类中心的程度，A(i,k)表示点x_i选择点x_k作为其聚类中心的适合程度；

步骤13：确定偏向参数$p_k=\underset{i\neq j,i,j=1,\cdots,N}{median}\ S(i,j)$，pk表示各样本数据点被选作聚类中心的可能性，是相似矩阵S对角线上元素的取值，k＝1，…，N，N表示输入节点的数量，median函数表示取一组数值中居于中间的数值；

步骤14：根据下述公式计算吸引度矩阵R和归属度矩阵A：

$$R(i,k)=S(i,k)-\max_{k\neq k'}\{A(i,k')+S(i,k')\}$$

$$A(i,k)=\min\left\{0,R(k,k)+\sum_{i'\notin\{i,k\}}\max\{0,R(i',k)\}\right\}$$

$$R(k,k)=p(k)-\max_{k\neq k'}\{A(k,k')+S(k,k')\}$$

其中p(k)表示数据点x_k作为聚类中心的参考度，R(k,k)表示数据点x_k适合作为自己的聚类中心的程度，A(k,k')表示数据点x_k选择数据点$x_{k'}$作为其聚类中心的程度，S(k,k')表示数据点x_k和数据点$x_{k'}$的相似程度；

步骤15：更新吸引度矩阵R和归属度矩阵A的公式为：

$R(i,k)=\lambda*R(i,k)_{old}+(1-\lambda)*R(i,k)_{new}$

$A(i,k)=\lambda*A(i,k)_{old}+(1-\lambda)*A(i,k)_{new}$

上述更新公式表示每次迭代时，新的吸引度矩阵$R(i,k)_{new}$和归属度矩阵$A(i,k)_{new}$要分别与上一次的吸引度矩阵$R(i,k)_{old}$和归属度矩阵$A(i,k)_{old}$进行加权更新，得到该次迭代的吸引度矩阵和归属度矩阵，其中λ表示更新因子；

步骤16：如果满足以下条件之一：①选择的类中心保持稳定，②超过最大迭代次数，则转至步骤17，否则转至步骤14；

步骤17,输出聚类结果。

3.根据权利要求1所述的基于APDE-RBF神经网络的网络安全态势预测方法,其特征在于,所述步骤2中进一步包括以下步骤:

步骤21:执行初始化,过程如下:

$\sigma_i=\sigma_{min}+rand(0,1)*(\sigma_{max}-\sigma_{min})$

$w_i=rand(0,1)$

其中σ_i为RBF神经网络基函数宽度,σ_{max}表示所有样本数据点中两个最远数据点的距离宽度,其计算公式为:$\sigma_{max}=\arg\max_{i\neq j,i,j=1,\ldots,h}(abs(c_i-c_j))$;$\sigma_{min}$表示所有样本数据点中两个最近数据点的距离宽度,其计算公式为:$\sigma_{min}=\arg\min_{i\neq j,i,j=1,\ldots,h}(abs(c_i-c_j))$;$c_i$、$c_j$表示任意两个不同的隐含层节点,$w_i$表示隐含层到输出层连接权值,rand(0,1)表示(0,1)间均匀分布的随机数;

步骤22:执行变异过程,将第g+1代种群中变异个体$V_i(g+1)$建模为第g代种群中三个个体的函数:

$V_i(g+1)=X_{r1}(g)+F*(X_{r2}(g)-X_{r3}(g))$

$i\neq r1\neq r2\neq r3$

其中$X_i(g)$是第g代种群中第i个个体,即$X_{r1}(g)$、$X_{r2}(g)$和$X_{r3}(g)$分别表示第g代种群中第r1个、第r2个以及第r3个个体,F为缩放因子;

步骤23:执行交叉过程,产生第g+1代第i个第j维新个体$u_{ij}(g+1)$的公式为:

$$u_{ij}(g+1)=\begin{cases}v_{ij}(g+1) & 当rand<CR\ 或者\ j=j_{rand}\\ x_{ij}(g) & 其它\end{cases}$$

其中$v_{ij}(g+1)$表示第g代种群第i个第j维个体进行变异操作后的个体,$x_{ij}(g)$表示第g代种群第i个第j维个体,rand是(0,1)间均匀分布的随机数,j_{rand}是[1,n]间的随机整数,CR表示交叉概率;上述公式含义为:当随机变量rand小于交叉概率CR或者个体中元素对应序数j等于随机变量j_{rand},则采用变异个体中的元素作为新个体,旨在提高个体变异的可能性;否则,仍保持目标个体$x_{ij}(g)$不变;

步骤24:执行选择过程,如下:

$$X_i(g+1)=\begin{cases}U_i(g+1) & 当f(U_i(g+1))\geq f(X_i(g))\\ X_i(g) & 其它\end{cases}$$

其中$U_i(g+1)$是候选个体,$X_i(g)$是对应个体,f(•)是个体的适应度函数,此处使用均方误差作为适应度函数。

4.根据权利要求3所述的基于APDE-RBF神经网络的网络安全态势预测方法,其特征在于,所述步骤22中对缩放因子F进行动态调整的公式为:

$$F(g)=\begin{cases}F_{\max}-(F_{\max}-F_{\min})*(\dfrac{g-1}{g_{\max}}) & 当PD(g)>PD(g-1)\ 且\ \dfrac{g}{g_{\max}}<\tau_1\\ F(g-1) & 当PD(g)\le PD(g-1)\ 且\ \dfrac{g}{g_{\max}}<\tau_1\\ F_{\min} & 其它\end{cases}$$

其中F_{max}和F_{min}分别表示缩放因子的上下界，PD(g)是第g代中的种群差异度，且种群差异度表示对种群空间中所有个体进行聚类所得到的聚类个数，当种群差异度越大时，个体在种群空间中分布越均匀，求得全局最优解可能性越大；τ_1为设置的迭代阈值，g_{max}为最大迭代次数。

5.根据权利要求3所述的基于APDE-RBF神经网络的网络安全态势预测方法，其特征在于，所述步骤23中的使交叉概率CR可自适应调整的公式为：

$$CR(g)=\begin{cases}CR_{\min}*2^{e^{(1-\frac{g_{\max}}{1+g})}} & 当\dfrac{g}{g_{\max}}<\tau_2\\ CR_{\max} & 其它\end{cases}$$

其中CR_{min}和CR_{max}分别表示交叉概率的上下界，τ_2为设置的迭代阈值，g_{max}为最大迭代次数。

6.根据权利要求1所述的基于APDE-RBF神经网络的网络安全态势预测方法，其特征在于，所述步骤3中混沌搜索具体为：首先建模一维Logistic映射混沌模型，其表达式为：$Z^{t+1}=\mu Z^t(1-Z^t)$，其中，Z^t是一个D维向量，μ是控制参数，t表示混沌迭代次数；其次，建模种群中最优个体和差异度中心迭代更新公式：

$$X_i^{t+1}=X_i+\alpha Z^{t+1}$$

$$\alpha=\begin{cases}1 & 当r\ge 0.5\\ -1 & 其它\end{cases}$$

其中X_i表示种群的最优个体或者差异度中心，X_i^{t+1}表示混沌搜索后的新个体，α表示混沌调节参数，r是[0,1]间的随机数。

基于APDE-RBF神经网络的网络安全态势预测方法

技术领域

[0001] 本发明属于网络安全技术领域，特别涉及一种基于吸引子传播差分进化算法的径向基函数(Affinity Propagation Differential evolution-Radial Basis Function，简称APDE-RBF)神经网络的网络安全态势预测方法。

背景技术

[0002] 根据2015年1月中国互联网信息中心发布的《第35次中国互联网发展状况报告》显示，截止2014年12月底，我国总体网民中有46.3%的网民遭遇过网络安全问题，表明我国个人互联网使用的安全状况不容乐观。随着网络安全问题日益突出与严重，一些传统的安全防御技术已力不从心，为解决上述问题，网络安全态势感知的研究应运而生。

[0003] 网络安全态势预测主要是在网络受到攻击损失前网络管理员采取相对应的措施，根据当前和以往的网络安全态势值，建立合理的数学模型对未来一段时间的网络安全状态进行预测。由于网络攻击是随机和不确定的，所以对态势值的预测是一个复杂的非线性过程。

[0004] 目前研究人员提出了很多预测的方法，如统计方法、灰色预测方法、神经网络方法、马尔科夫模型、支持向量机等，但上述方法都存在各自的局限性与不足。

[0005] 统计方法较为常用的模型有：自回归模型、滑动平均模型和自回归滑动平均模型，然而这些模型存在以下局限性：时间序列的数据要求平稳，如果是多元回归，还要求变量之间是独立的；灰色预测方法适用单调变化的时间序列，对于波动较大的时间序列难以预测；1987年，Lapdes等人首先将神经网络应用于由计算机产生的时间序列仿真数据的学习和预测，但神经网络存在收敛速度慢、结构选择困难和容易陷入局部极小等问题，同时由于该方法受网络结构复杂度和样本复杂度的影响较大，因而会出现过学习或泛化能力低的现象；马尔科夫模型需要大量复杂的数学公式推导，难以建立准确的预测模型；支持向量机(Support Vector Machine，简称SVM)对大规模训练样本难以实施，收敛速度慢。

发明内容

[0006] 针对上述现有技术的不足，本发明提出一种基于APDE-RBF神经网络的网络安全态势预测方法，旨在增强泛化能力的同时，提高对网络安全态势的预测精度。

[0007] 为实现上述目的，本发明提出的一种基于APDE-RBF神经网络的网络安全态势预测方法，其特征在于，该预测方法包括以下步骤：

[0008] 步骤1：利用吸引力传播(Affinity Propagation，简称AP)聚类算法对样本数据进行划分聚类，从而获得RBF的中心和网络的隐含层节点数；

[0009] 步骤2：利用AP聚类得出种群差异度，自适应地改变差分进化(Differential evolution，简称DE)算法的缩放因子和交叉概率，对RBF的宽度和连接权值进行优化；

[0010] 步骤3：为了避免陷入局部最优以及跳出局部极值点，对每一代种群的精英个体和种群差异度中心进行混沌搜索；

[0011] 步骤4:确定最终RBF网络模型,输入测试数据集,输出态势预测值。

[0012] 优选地,所述步骤1进一步包括以下步骤:

[0013] 步骤11:利用欧氏距离计算输入节点之间的相似度矩阵S为:S(i,k)=-||x_i-x_k||^2,其中x_i和x_k表示RBF神经网络任意两个输入节点,S(i,k)表示点x_k作为点x_i的聚类中心的相似度;

[0014] 步骤12:初始化吸引度矩阵R和归属度矩阵A为R(i,k)=0,A(i,k)=0,其中R(i,k)表示点x_k适合作为数据点x_i的聚类中心的程度,A(i,k)表示点x_i选择点x_k作为其聚类中心的适合程度;

[0015] 步骤13:确定偏向参数$p_k=\underset{i\neq j,i,j=1,\cdots,N}{median}\ S(i,j)$,其中N表示输入节点的数量,median函数表示一组数值中居于中间的数值;

[0016] 步骤14:根据下述公式计算吸引度矩阵R和归属度矩阵A:

[0017] $$R(i,k)=S(i,k)-\max_{k\neq k'}\{A(i,k')+S(i,k')\}$$

[0018] $$A(i,k)=\min\left\{0,R(k,k)+\sum_{i'\notin\{i,k\}}\max\{0,R(i',k)\}\right\}$$

[0019] $$R(k,k)=p(k)-\max_{k\neq k'}\{A(k,k')+S(k,k')\}$$

[0020] 其中p(k)表示数据点x_k作为聚类中心的参考度,R(k,k)表示数据点x_k适合作为自己的聚类中心的程度,A(k,k')表示数据点x_k选择数据点$x_{k'}$作为其聚类中心的程度,S(k,k')表示数据点x_k和数据点$x_{k'}$的相似程度;从上述公式可看出当p(k)较大时其对应的R(k,k)也会较大,进而A(i,k)取值也会较大,从而类代表k作为最终聚类中心的可能性较大;相应地,当越多的p(k)较大时,越多的类代表倾向于成为最终的聚类中心,因此,增大或者减小p(k)可以增加或减少AP输出的聚类数目;

[0021] 步骤15:更新吸引度矩阵R和归属度矩阵A的公式为:

[0022] $R(i,k)=\lambda*R(i,k)_{old}+(1-\lambda)*R(i,k)_{new}$

[0023] $A(i,k)=\lambda*A(i,k)_{old}+(1-\lambda)*A(i,k)_{new}$

[0024] 上述更新公式表示每次迭代时,新的吸引度矩阵$R(i,k)_{new}$和归属度矩阵$A(i,k)_{new}$要分别与上一次的$R(i,k)_{old}$和$A(i,k)_{old}$进行加权更新,得到该次迭代的吸引度矩阵和归属度矩阵,其中λ表示更新因子;

[0025] 步骤16:如果满足以下条件之一:①选择的类中心保持稳定,②超过最大迭代次数,则转至步骤17,否则转步骤14;

[0026] 步骤17,输出聚类结果。

[0027] 优选地,所述步骤2进一步包括以下步骤:

[0028] 步骤21:执行初始化,过程如下:

[0029] $\sigma_i=\sigma_{min}+rand(0,1)*(\sigma_{max}-\sigma_{min})$

[0030] $w_i=rand(0,1)$

[0031] 其中σ_i为RBF神经网络基函数宽度,σ_{max}表示所有样本数据点中两个最远数据点的距离宽度,其计算公式为:$\sigma_{max}=\underset{i\neq j,i,j=1,\ldots,h}{\arg\max}(abs(c_i-c_j))$;$\sigma_{min}$表示所有样本数据点中两个最

近数据点的距离宽度，其计算公式为：$\sigma_{\min}=\underset{i\neq j,i,j=1,\ldots,h}{\arg\min}(abs(c_i-c_j))$；$w_i$表示隐含层到输出层连接权值，rand(0,1)表示(0,1)间均匀分布的随机数；

[0032] 步骤22：执行变异过程，将第g+1代种群中变异个体$V_i(g+1)$建模为第g代种群中三个个体的函数：

[0033] $V_i(g+1)=X_{r1}(g)+F*(X_{r2}(g)-X_{r3}(g))$

[0034] $i\neq r1\neq r2\neq r3$

[0035] 其中$X_i(g)$是第g代种群中第i个个体，即$X_{r1}(g)$、$X_{r2}(g)$和$X_{r3}(g)$分别表示第g代种群中第r1个、第r2个以及第r3个个体，F为缩放因子；

[0036] 步骤23：执行交叉过程，产生第g+1代第i个第j维新个体$u_{ij}(g+1)$的公式为：

[0037] $$u_{ij}(g+1)=\begin{cases}v_{ij}(g+1) & rand<CR\ or\ j=j_{rand}\\ x_{ij}(g) & otherwise\end{cases}$$

[0038] 其中$v_{ij}(g+1)$表示第g代种群第i个第j维个体进行变异操作后的个体，$x_{ij}(g)$表示第g代种群第i个第j维个体，rand是(0,1)间均匀分布的随机数，j_{rand}是[1,n]间的随机整数，CR表示交叉概率；上述公式含义为当随机变量rand小于交叉概率CR或者个体中元素对应序数j等于随机变量j_{rand}，即采用变异个体中的元素作为新个体，旨在提高个体变异的可能性；否则，仍保持目标个体$x_{ij}(g)$不变；

[0039] 步骤24：执行选择过程，如下：

[0040] $$X_i(g+1)=\begin{cases}U_i(g+1) & f(U_i(g+1))\geq f(X_i(g))\\ X_i(g) & otherwise\end{cases}$$

[0041] 其中$U_i(g+1)$是候选个体，$X_i(g)$是对应个体，f(•)是个体的适应度函数，此处使用均方误差(mean square error，简称MSE)作为适应度函数。

[0042] 进一步，所述步骤22中对缩放因子F进行动态调整的公式为：

[0043] $$F(g)=\begin{cases}F_{\max}-(F_{\max}-F_{\min})*(\frac{g-1}{g_{\max}}) & PD(g)>PD(g-1)\ and\ \frac{g}{g_{\max}}<\tau_1\\ F(g-1) & PD(g)\leq PD(g-1)\ and\ \frac{g}{g_{\max}}<\tau_1\\ F_{\min} & otherwise\end{cases}$$

[0044] 其中F_{max}和F_{min}分别表示缩放因子的上下界，PD(g)是第g代中的种群差异度，且种群差异度表示对种群空间中所有个体进行聚类所得到的聚类个数，当种群差异度越大时，个体在种群空间中分布越均匀，求得全局最优解可能性越大；τ_1为迭代阈值，g_{max}为最大迭代次数。

[0045] 进一步，所述步骤23中的使交叉概率CR可自适应调整的公式为：

[0046] $$CR(g)=\begin{cases} CR_{\min}\cdot 2^{e^{\left(1-\frac{g_{\max}}{1+g}\right)}} & \frac{g}{g_{\max}}<\tau_2 \\ CR_{\max} & otherwise \end{cases}$$

[0047] 其中CR_{min}和CR_{max}分别表示交叉概率的上下界，τ_2为设置的迭代阈值。

[0048] 优选地，所述步骤3中混沌搜索具体实现为：首先建模一维Logistic映射混沌模型，其表达式为：$Z^{t+1}=\mu Z^t(1-Z^t)$，其中μ是控制参数，$Z_i^0=[z_1^0,z_2^0,...,z_D^0]$是一个随机生成的D维向量，t表示混沌迭代次数；

[0049] 其次，建模种群中最优个体和差异度中心迭代更新公式：

[0050] $$X_i^{t+1}=X_i+\alpha Z^{t+1}$$

[0051] $$\alpha=\begin{cases} 1 & r\geq 0.5 \\ -1 & otherwise \end{cases}$$

[0052] 其中X_i表示种群的最优个体或者差异度中心，X_i^{t+1}表示混沌搜索后的新个体，α表示混沌调节参数，r是[0,1]间的随机数。

[0053] 本发明的有益效果在于：采用本发明提出的AP聚类算法得出种群差异度，并自适应地改变DE算法的缩放因子和交叉概率，不仅优化了RBF的宽度和连接权值，而且对每一代种群的精英个体和种群差异度中心均进行混沌搜索，避免了陷入局部最优，不但能够增强泛化能力，而且能够提高预测精度。

附图说明

[0054] 图1是本发明提供的基于APDE-RBF神经网络网络安全态势预测方法的优选实施例流程图；

[0055] 图2是不同算法态势值预测对比仿真图；

[0056] 图3是不同算法的不同误差对比仿真图；

[0057] 图4是不同改进DE算法预测对比仿真图；

[0058] 图5是不同改进DE算法的不同误差对比仿真图。

具体实施方式

[0059] 为使本发明的目的、技术方案和优点更加清楚明白，下面结合附图对本发明的具体实施方式作进一步详细说明。

[0060] 图1是本发明提供的基于APDE-RBF神经网络的网络安全态势预测方法的优选实施例流程图，该方法具体包括以下步骤：

[0061] 步骤1：利用AP聚类算法对样本数据进行划分聚类，从而获得RBF的中心和网络的隐含层节点数；

[0062] 步骤2：利用AP聚类得出种群差异度，自适应地改变DE算法的缩放因子和交叉概率，对径向基函数RBF的宽度和连接权值进行优化；

[0063] 步骤3：为了避免陷入局部最优以及跳出局部极值点，对每一代种群的精英个体和种群差异度中心进行混沌搜索；

[0064] 步骤4:确定最终RBF网络模型,输入测试数据集,输出态势预测值。

[0065] 根据本发明,所述步骤1中进一步包括以下步骤:

[0066] 步骤11:利用欧氏距离计算输入节点之间的相似度矩阵S为:S(i,k)=-||x_i-x_k||^2,其中x_i和x_k表示RBF神经网络任意两个输入节点,S(i,k)表示点x_k作为点x_i的聚类中心的相似度;

[0067] 步骤12:初始化吸引度矩阵R和归属度矩阵A为R(i,k)=0,A(i,k)=0,其中R(i,k)表示点x_k适合作为数据点x_i的聚类中心的程度,A(i,k)表示点x_i选择点x_k作为其聚类中心的适合程度;

[0068] 步骤13:确定偏向参数$p_k=\underset{i\neq j,i,j=1,\cdots,N}{median}\ S(i,j)$,$p_k$表示各样本数据点被选作聚类中心的可能性,是相似矩阵S对角线上元素的取值,k=1,…,N,N表示输入节点的数量,median函数表示一组数值中居于中间的数值;

[0069] 步骤14:根据下述公式计算吸引度矩阵R和归属度矩阵A:

[0070] $$R(i,k)=S(i,k)-\max_{k\neq k'}\{A(i,k')+S(i,k')\}$$

[0071] $$A(i,k)=\min\left\{0,R(k,k)+\sum_{i'\notin\{i,k\}}\max\{0,R(i',k)\}\right\}$$

[0072] $$R(k,k)=p(k)-\max_{k\neq k'}\{A(k,k')+S(k,k')\}$$

[0073] 其中p(k)表示数据点x_k作为聚类中心的参考度,R(k,k)表示数据点x_k适合作为自己的聚类中心的程度,A(k,k')表示数据点x_k选择数据点x_k'作为其聚类中心的程度,S(k,k')表示数据点x_k和数据点x_k'的相似程度;从上述公式可看出当p(k)较大时其对应的R(k,k)也会较大,进而A(i,k)取值也会较大,从而类代表k作为最终聚类中心的可能性较大;相应地,当越多的p(k)较大时,越多的类代表倾向于成为最终的聚类中心,因此,增大或者减小p(k)可以增加或减少AP输出的聚类数目;

[0074] 步骤15:更新吸引度矩阵R和归属度矩阵A的公式为:

[0075] $R(i,k)=\lambda*R(i,k)_{old}+(1-\lambda)*R(i,k)_{new}$

[0076] $A(i,k)=\lambda*A(i,k)_{old}+(1-\lambda)*A(i,k)_{new}$

[0077] 上述更新公式表示每次迭代时,新的吸引度矩阵$R(i,k)_{new}$和归属度矩阵$A(i,k)_{new}$要分别与上一次的$R(i,k)_{old}$和$A(i,k)_{old}$进行加权更新,得到该次迭代的吸引度矩阵和归属度矩阵,其中λ表示更新因子;

[0078] 步骤16:如果满足以下条件之一:①选择的类中心保持稳定,②超过最大迭代次数,则转至步骤17,否则转步骤14;

[0079] 步骤17,输出聚类结果。

[0080] 根据本发明,所述步骤2中进一步包括以下步骤:

[0081] 步骤21:执行初始化,过程如下:

[0082] $\sigma_i=\sigma_{min}+rand(0,1)*(\sigma_{max}-\sigma_{min})$

[0083] $w_i=rand(0,1)$

[0084] 其中σ_i为RBF神经网络基函数宽度,σ_{max}表示所有样本数据点中两个最远数据点的

距离宽度，其计算公式为：$\sigma_{\max}=\underset{i\neq j,i,j=1,\ldots,h}{\arg\max}(abs(c_i-c_j))$；$\sigma_{\min}$表示所有样本数据点中两个最近数据点的距离宽度，其计算公式为：$\sigma_{\min}=\underset{i\neq j,i,j=1,\ldots,h}{\arg\min}(abs(c_i-c_j))$；$c_i$、$c_j$表示任意两个不同的隐含层节点，$w_i$表示隐含层到输出层连接权值，rand(0,1)表示(0,1)间均匀分布的随机数；

[0085] 步骤22：执行变异过程，将第g+1代种群中变异个体$V_i(g+1)$建模为第g代种群中三个个体的函数：

[0086] $V_i(g+1)=X_{r1}(g)+F*(X_{r2}(g)-X_{r3}(g))\ i\neq r1\neq r2\neq r3$

[0087] 其中$X_{r1}(g)$、$X_{r2}(g)$和$X_{r3}(g)$分别表示第g代种群中第r1个、第r2个以及第r3个个体，F为缩放因子；

[0088] 步骤23：执行交叉过程，产生新个体$u_{ij}(g+1)$的公式如下所示：

[0089] $$u_{ij}(g+1)=\begin{cases}v_{ij}(g+1) & 当rand<CR\ 或者\ j=j_{rand}\\ x_{ij}(g) & 其它\end{cases}$$

[0090] 其中$v_{ij}(g+1)$表示第g代种群第i个第j维个体进行变异操作后的个体，$x_{ij}(g)$表示第g代种群第i个第j维个体，rand是(0,1)间均匀分布的随机数，j_{rand}是[1,n]间的随机整数，CR表示交叉概率；上述公式含义为当随机变量rand小于交叉概率CR或者个体中元素对应序数j等于随机变量j_{rand}，即采用变异个体中的元素作为新个体，旨在提高个体变异的可能性；否则，仍保持目标个体$x_{ij}(g)$不变；

[0091] 步骤24：执行选择过程，具体如下：

[0092] $$X_i(g+1)=\begin{cases}U_i(g+1) & 当f(U_i(g+1))\geq f(X_i(g))\\ X_i(g) & 其它\end{cases}$$

[0093] 其中$U_i(g+1)$是候选个体，$X_i(g)$是对应个体，f(•)是个体的适应度函数，此处使用均方误差(mean square error，简称MSE)作为适应度函数。

[0094] 根据本发明，所述步骤22中的对F进行动态调整的公式为：

[0095] $$F(g)=\begin{cases}F_{\max}-(F_{\max}-F_{\min})*(\frac{g-1}{g_{\max}}) & 当PD(g)>PD(g-1)\ 且\ \frac{g}{g_{\max}}<\tau_1\\ F(g-1) & 当PD(g)\leq PD(g-1)\ 且\ \frac{g}{g_{\max}}<\tau_1\\ F_{\min} & 其它\end{cases}$$

[0096] 其中$F_{\max}$和$F_{\min}$分别表示缩放因子的上下界，PD(g)是第g代中的种群差异度，且种群差异度表示对种群空间中所有个体进行聚类所得到的聚类个数，当种群差异度越大时，个体在种群空间中分布越均匀，求得全局最优解可能性越大；τ_1为设置的迭代阈值，$g_{\max}$为最大迭代次数。

[0097] 根据本发明，进一步，所述步骤23中的使交叉概率CR可自适应调整的公式为：

[0098] $$CR(g)=\begin{cases}CR_{\min}*2^{e^{\left(1-\frac{g_{\max}}{1+g}\right)}} & 当\dfrac{g}{g_{\max}}<\tau_2\\ CR_{\max} & 其它\end{cases}$$

[0099] 其中CR_{min}和CR_{max}是交叉概率的上下界，τ_2是设置的迭代阈值。

[0100] 根据本发明，进一步，所述步骤3中混沌搜索具体实现为：首先建模一维Logistic映射混沌模型，其表达式为：$Z^{t+1}=\mu Z^t(1-Z^t)$该式是数学意义上的迭代公式，其值$Z_i^0=[z_1^0,z_2^0,...,z_D^0]$是一个随机生成的D维向量，其中，μ是控制参数，t表示混沌迭代次数；

[0101] 其次，建模种群中最优个体和差异度中心迭代更新公式：

[0102] $$X_i^{t+1}=X_i+\alpha Z^{t+1}$$

[0103] $$\alpha=\begin{cases}1 & 当r\geq 0.5\\ -1 & 其它\end{cases}$$

[0104] 其中X_i表示种群的最优个体或者差异度中心，X_i^{t+1}表示混沌搜索后的新个体，α表示混沌调节参数，r是[0,1]间的随机数。

[0105] 为说明本发明的有益效果，根据仿真结果进行进一步分析。

[0106] 图2描述了基于不同算法得到的网络安全态势值。从图2可以看出，自回归滑动平均模型(Auto-Regressive and Moving Average Model，简称ARMA)主要针对随机平稳的时间序列，但是因为网络攻击的随机性和复杂性，网络安全态势序列是非平稳的；灰色模型(Grey Model，简称GM)对于单调变化的时间序列预测效果好，反之误差大；最小二乘支持向量机(Least Squares Support Vector Machines，简称LSSVM)的支持向量变成了所有数据点，失去了SVM的稀疏性特点；Kmeans-RBF需要预先设定隐含层节点，忽略了数据本身的特点，弱化了RBF的泛化能力，然而，以真实值作为衡量标准，相比于上述方法出现的不同程度的误差和缺陷，本发明提出的APDE-RBF神经网络模型预测精度最高。

[0107] 图3显示了不同算法的不同误差对比，从图3可以看出不论是平均相对误差、均方根误差还是相对均方误差，APDE-RBF神经网络模型都保持在较小的误差水平，体现了较高的预测精度。

[0108] 图4显示不同改进DE算法在不同时间点的网络安全态势值。DE算法是固定的F和CR，易陷入局部最优；简化的差分进化版本(Simplified Differential Evolution Version，简称SDE)算法的F采用简单的随机数；基于全体参数和变异策略的差分进化算法(Differential evolution algorithm with ensemble of parameters and mutation strategies，简称EPSDE)算法利用变异策略池和参数池随机组合进行迭代进化；自适应差分进化(Self-Adaptive Differential Evolution，简称jDE)算法的F和CR依赖随机数判别从而得到不同的结果；基于复合试验向量生成策略和控制参数的差分进化(Evolution With Composite Trial Vector Generation Strategies and Control Parameters，简称CoDE)算法是利用三种不同的变异策略和参数设置竞争耦合进行迭代进化。上述方法虽然对DE算法的变异策略和参数设置进行自适应改进，但是大多都是随机数或依赖随机数进行判别选取，导致进化不稳定。进一步，从图中可以看出，以真实值作为衡量标准，相比于上述

算法，APDE算法总体上维持了较低的绝对误差，其原因在于APDE-RBF神经网络模型依赖种群差异度和迭代进化程度对F和CR进行自适应调整，使种群向有利方向进化，加快了算法的收敛速度。

[0109] 图5是不同算法的不同误差对比，从图5可以看出不论是平均相对误差、均方根误差还是相对均方误差，本发明提出的APDE-RBF神经网络模型都保持在较小的误差水平，体现了较高的预测精度。

[0110] 本发明所举实施方式或者实施例对本发明的目的、技术方案和优点进行了进一步的详细说明，所应理解的是，以上所举实施方式或者实施例仅为本发明的优选实施方式而已，并不用以限制本发明，凡在本发明的精神和原则之内对本发明所作的任何修改、等同替换、改进等，均应包含在本发明的保护范围之内。

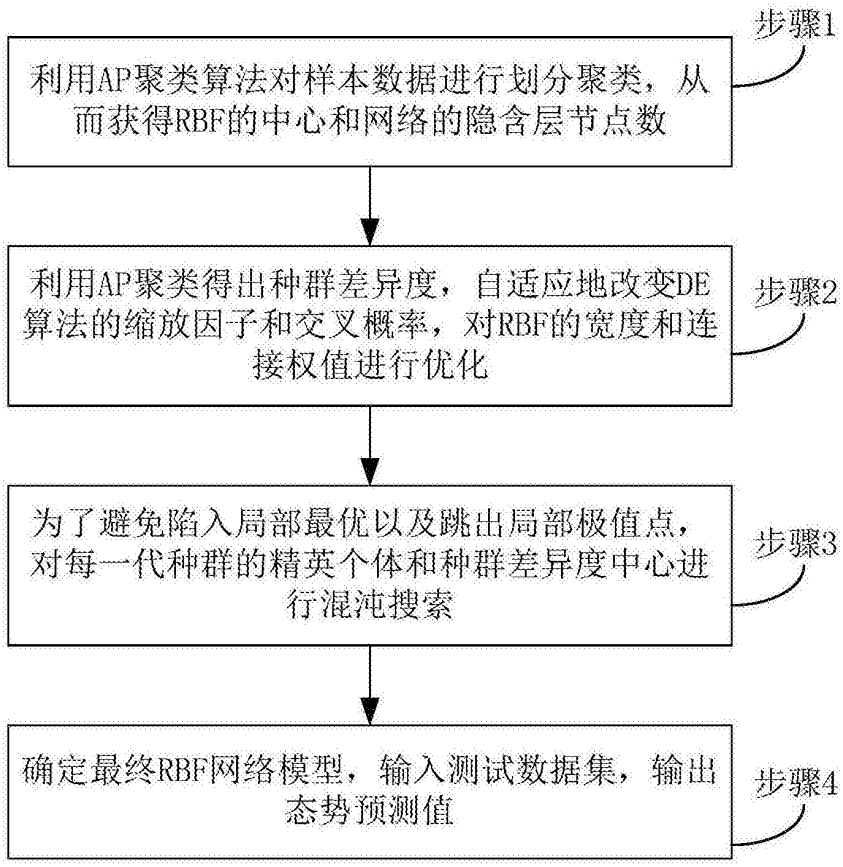

步骤1
利用AP聚类算法对样本数据进行划分聚类，从而获得RBF的中心和网络的隐含层节点数
利用AP聚类得出种群差异度，自适应地改变DE算法的缩放因子和交叉概率，对RBF的宽度和连接权值进行优化
步骤2
为了避免陷入局部最优以及跳出局部极值点，对每一代种群的精英个体和种群差异度中心进行混沌搜索
步骤3
确定最终RBF网络模型，输入测试数据集，输出态势预测值
步骤4

图1

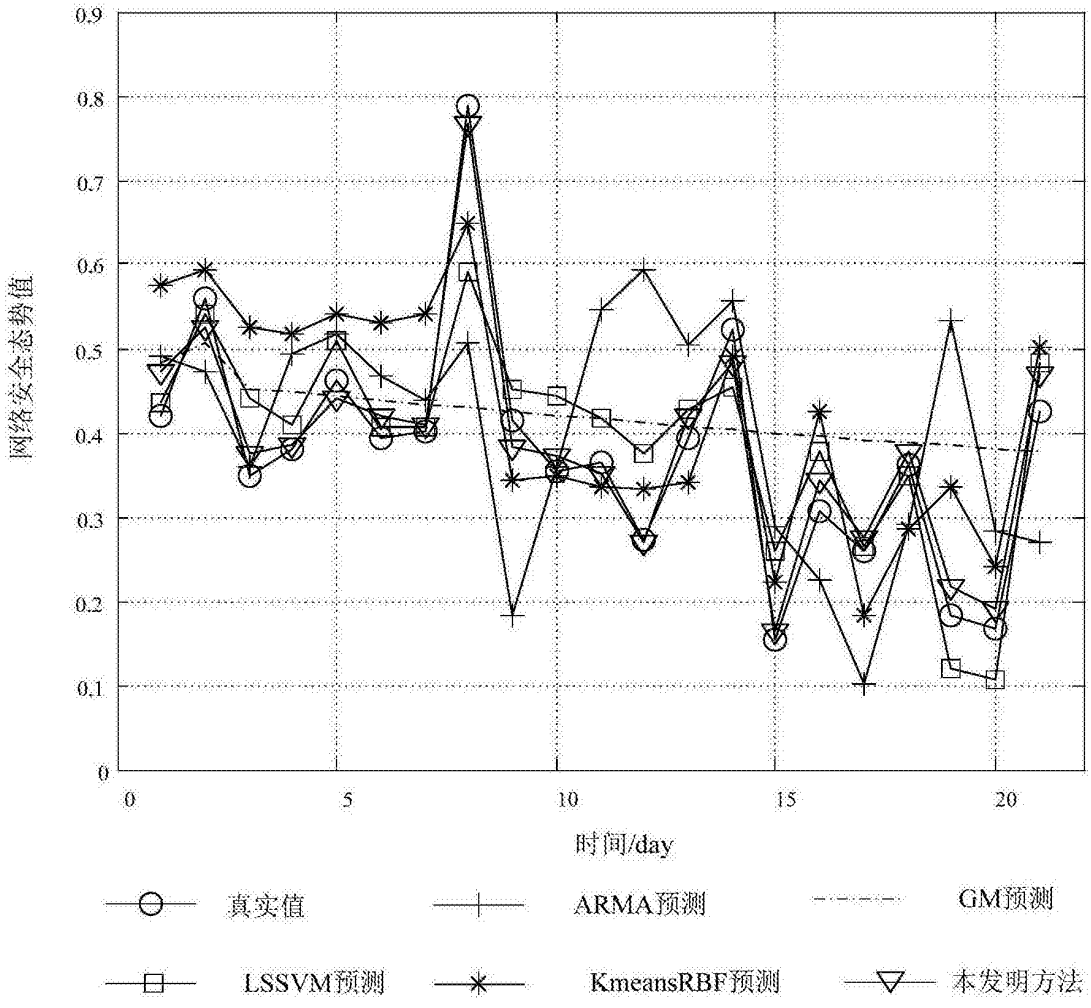

网络安全态势值
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
5
10
15
20
时间/day
真实值
ARMA预测
GM预测
LSSVM预测
KmeansRBF预测
本发明方法

图2

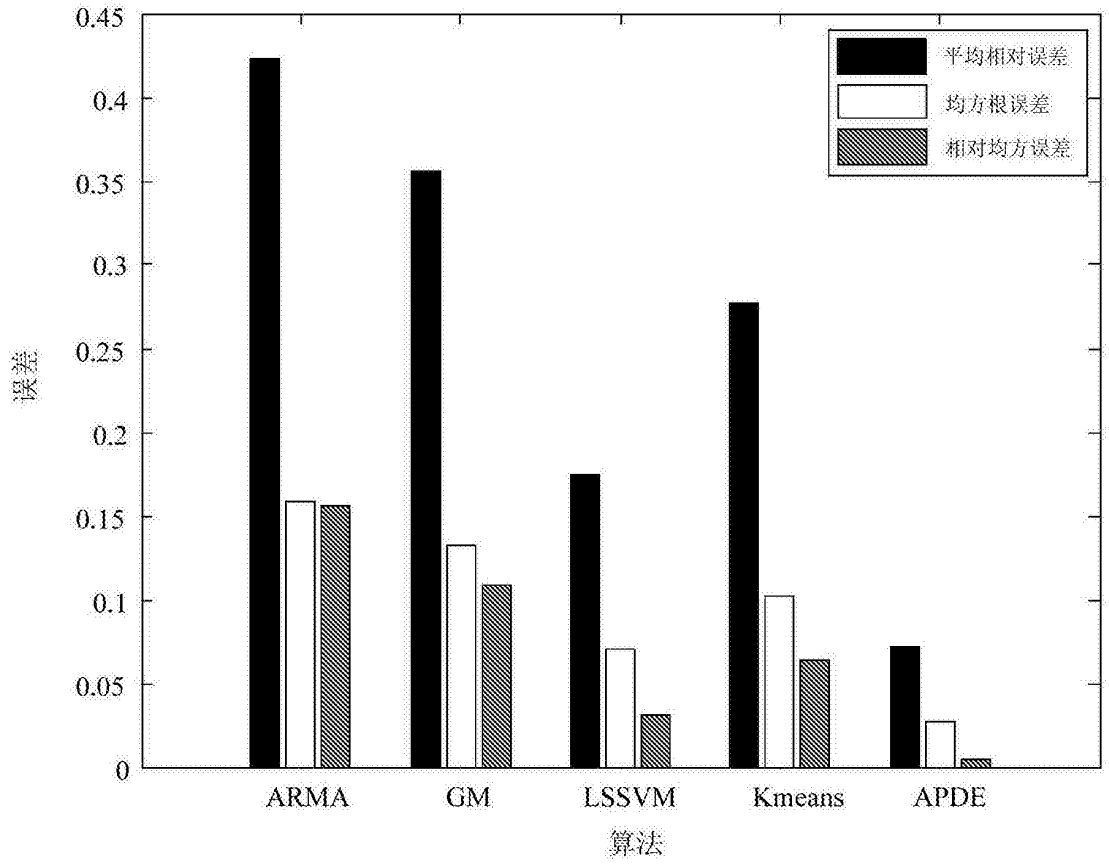

0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
误差
平均相对误差
均方根误差
相对均方误差
ARMA
GM
LSSVM
Kmeans
APDE
算法

图3

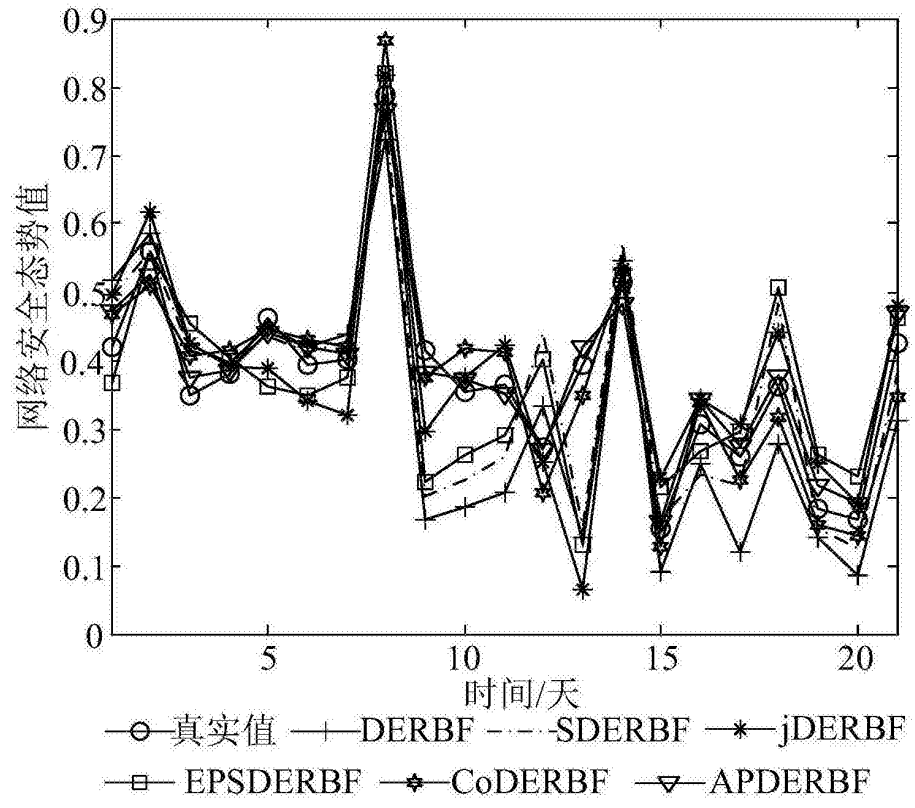

网络安全态势值
时间/天
真实值
DERBF
SDERBF
jDERBF
EPSDERBF
CoDERBF
APDERBF

图4

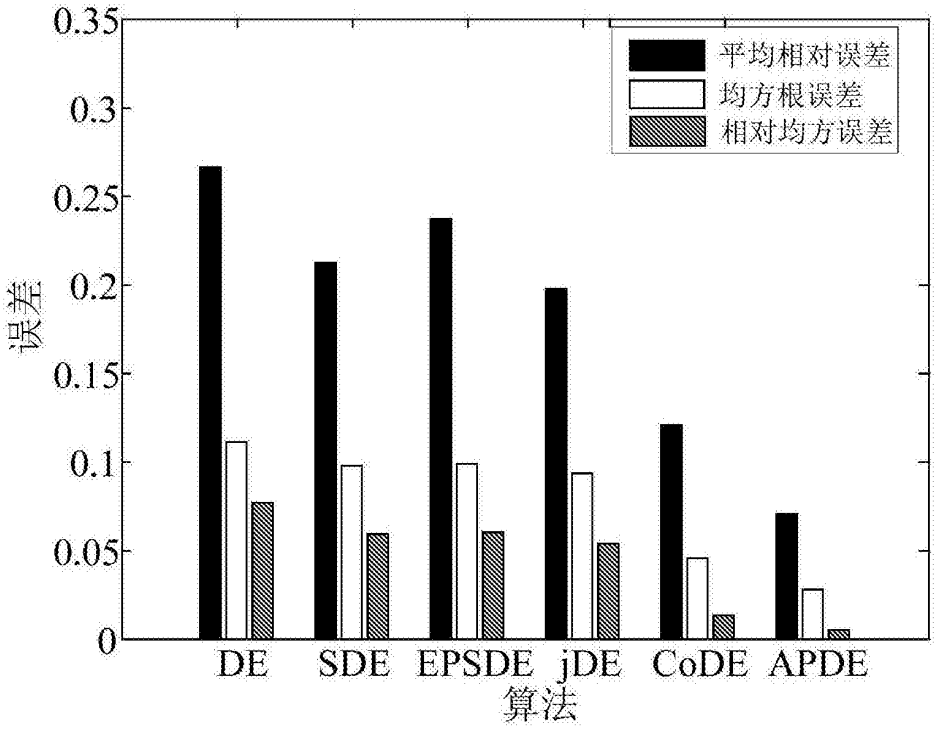

平均相对误差
均方根误差
相对均方误差
误差
DE
SDE
EPSDE
jDE
CoDE
APDE
算法

图5