

# 人工免疫网络算法与RBF网结合的应用研究

颜学明 杨凌霄 葛红

(华南师范大学计算机学院 广州 510631)

**摘要** 人工免疫网络是一种基于生物免疫系统启示的智能信息处理模型,得到广泛的研究和应用。在改进现有的人工免疫网络算法的基础上,将其与RBF网络结合构成基于人工免疫网络的RBF网并应用于问题求解。实验表明,改进后的人工免疫网络算法具有减少数据冗余、平滑噪声、提取特征数据从而提高数据质量的功能,在与RBF网结合的应用中可以有效提高问题求解的速度和性能。

**关键词** 人工免疫网络,特征数据,RBF网,函数逼近

中图分类号 TP183 文献标识码 A

## Application Research on Artificial Immune Network Combined with RBF

YAN Xue-ming YANG Ling-xiao GE Hong

(School of Computer Science, South China Normal University, Guangzhou 510631, China)

**Abstract** Artificial immune network, an intelligent information processing model inspired by biological immune system, has been widely studied and applied. In this paper, a modified artificial immune algorithm combined with traditional RBF was used in problem solving. The experimental results show that the modified artificial immune network algorithm can present the functions of reducing data redundancy, smoothing noise and extracting feature data to improve data quality. In the application of combining with RBF network, the speed and functionality of problem solving are raised effectively.

**Keywords** Artificial immune network, Feature data, RBF network, Function approximation

## 1 概述

人工免疫网络模型是通过借鉴和利用生物免疫系统的功能和机制开发,用于解决工程和科学问题的智能信息处理技术<sup>[1]</sup>。在Jerne免疫网络理论的启示下,de Castro和Timmis<sup>[2,3]</sup>两人分别提出了用于数据压缩和聚类的免疫网络智能信息处理模型,其中,aiNet(Artificial Immune NETwork)模型用一个小规模的“内镜像”记忆网络映射原输入数据集,从而达到数据压缩和数据特征提取的目的。从工作原理上,它属于非线性特征提取方法中的“仿生模式识别”技术,注重保留原始数据空间分布中的拓扑结构,比如数据固有聚类的空间分布等。由于最终结果是原始数据集的特征数据,而不是其子集,因此,它不同于通常的数据压缩技术和模式识别中的特征选择技术;免疫网络模型的这种数据提取方法具有较强的容错性能和联想能力,不会因为部分输入信号受损而严重影响其总体性能<sup>[4,5]</sup>;也不会因为输入信号受到一定程度噪声的污染而严重歪曲输出。

如前所述,由于人工免疫网络算法主要适用特征数据提取,属于数据预处理的过程,因此通常需要与其他算法结合才能解决工程应用问题。

RBF网络是一种三层前向神经网络模型,它具有全局逼近特性、计算量小、学习速度快等优点,无论在逼近能力、收敛

速度、搜索的遍历性和重构的精度等方面均优于其它前向网络,目前已经广泛应用于模式分类、系统辨识、函数逼近等领域。在RBF网络中,隐层中心的数量和位置的选择是整个网络性能优劣的关键,直接影响网络的逼近能力。因此RBF网络研究的重点和难点就在于隐层节点数的确定和隐层中心的确定。目前对于隐层中心的确定大多采用K均值聚类算法。该算法需要预先指定隐层中心的数量K,这是一件非常困难的事情。由于数据集中常常包含了采样过程的噪声与误差,这往往掩盖了原本数据集中的主要内容,或者使得提取有用的信息更加困难,因此利用K均值算法得到隐层的位置并不能很好地代表输入数据的主要特征<sup>[6,7]</sup>。

因此,本文提出改进的人工免疫网络算法(Modified Artificial Immune Network,简称MFAIN)用于提取原数据集的特征数据并以该特征数据作为RBF网隐层节点的中心值,从而自适应地确定聚类中心的数量和位置,然后利用LMS算法求出输出层的权值。

接下来本文首先对现有的人工免疫网络存在的问题做出几点改进,并对算法MFAIN做了详细地描述;然后对基于MFAIN算法的RBF网做出介绍;最后通过仿真实验,证明了所提方法的可行性和有效性。

## 2 人工免疫网络的改进

目前的人工免疫网络算法并不能很好地实现抗体多样性

本文受基金项目“在线空间叠合分析理论与方法研究”(60842007)资助。

颜学明(1987—),女,硕士生,主要研究方向为智能信息处理,E-mail: xueming126@126.com;杨凌霄(1988—),男,硕士生,主要研究方向为智能信息处理;葛红(1969—),女,博士,副教授,主要研究方向为智能计算方法及应用等。

和避免未成熟收敛问题<sup>[7-9]</sup>,因此本文对传统的免疫网络模型做出了改进,即采用抗体生存期望值对抗体进行抑制,对亲和力高的抗体进行促进,对高浓度的抗体进行抑制;同时本文对网络抑制采取的方法也有不同,选取的是被抗原激活频率高的抗体作为最后的网络输出,这样能更好提取抗原数据中的有用的特征信息,从而具有一定的平滑和抗噪性能。

## 2.1 MFAIN 算法描述

MFAIN 算法的流程图如图 1 所示,本文提出的 MFAIN 算法采用实数编码,抗原  $A_g$  就是原始数据集,输出的抗体集  $Ab$  就是提取出来的特征数据。

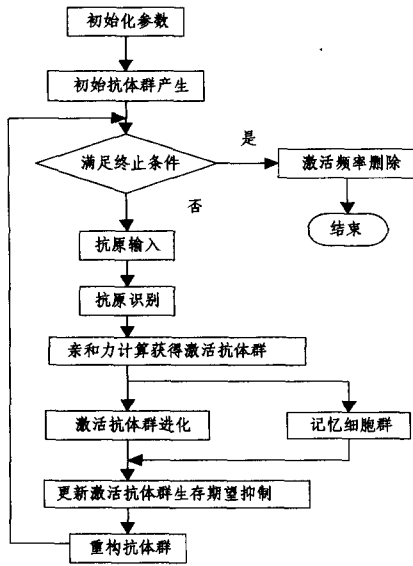


图 1 MFAIN 算法流程图

### 2.1.1 初始抗体的产生

在一定范围内随机生成抗体群,并将每一个抗体的激活标志都设置为 0。本算法采用实数编码,每个抗体实际上是一个实数向量。

### 2.1.2 抗原识别

抗原识别取决于抗原与抗体之间的亲和力,其公式定义如下。

$$A_{ij} = \frac{1}{1 + \|A_{g_i} - Ab_{j_i}\|} \quad (1)$$

式中,  $\|A_{g_i} - Ab_{j_i}\|$  为  $A_{g_i}$  与  $Ab_{j_i}$  之间的欧氏距离。而抗体与抗体之间的亲和力直接用欧氏距离表示,抗体与抗体之间的距离越小,则亲和力越小,抗体之间的相似度越大。

### 2.1.3 激活抗体群进化

#### 1) 克隆

根据克隆选择的思想,与抗原亲和力较高的抗体要进行克隆操作,亲和力越大则克隆的个数越多。若选取亲和力较高的前  $L$  个进行克隆,克隆的总规模数为  $S$ ,则该抗体  $Ab_{j_i}$  克隆的个数为

$$Q_j = \text{Int}(S * \frac{A_{ij}}{\sum_{i=1}^L A_{ij}}) \quad (2)$$

式中,  $\text{Int}(x)$  为大于  $x$  的最小整数。

#### 2) 超变异

对克隆后的细胞还需要超变异,变异的公式如下。

$$Ab_{k'} = Ab_k + \beta_k (A_{g_i} - Ab_k) \quad (3)$$

式中,  $\beta_k$  为变异率,且

$$\beta_k = e^{1 - \frac{S+1}{S}} \quad (4)$$

式中,  $\text{rand}$  为 0 到 1 的随机数。

### 2.1.4 抗体浓度的定义以及生存期望的计算

在免疫系统中,当一种抗体受到抗原刺激或其他抗体刺激或抑制时,这种抗体的数量将发生变化,亲和力高的抗体浓度提高,高到一定值要受到抑制,反之相应提高浓度低的抗体的产生和选择概率,在由  $N$  个组成的抗体集合中,抗体的浓度计算公式如下。

$$c_i = \frac{\sum_{i=1}^N S_{ij}}{N} \quad (5)$$

$$S_{ij} = \begin{cases} 1, & Ab_{ij} / \max\{Ab_{ij}\} \geq \eta \\ 0, & Ab_{ij} / \max\{Ab_{ij}\} < \eta \end{cases} \quad (6)$$

式中,  $Ab_{ij}$  为抗体  $Ab_i$  与抗体  $Ab_j$  之间的亲和力,  $\eta$  为抗体之间的相似系数。同时抗体的  $Ab_i$  生存期望定义为

$$E_i = \frac{A_{ij}}{C_i} \quad (7)$$

### 2.1.5 激活频率删除

保存那些被抗原多次激活的抗体,这样能更好地获得代表抗原集的特征,而对那些激活频率低或者从未被激活的抗体予以删除。设抗体  $Ab_i$  被抗原激活的次数为  $n_i$ ,若

$$n_i < \frac{\sum_{i=1}^N n_i}{N}, \text{ 或 } n_i = 0 \quad (8)$$

则将抗体  $Ab_i$  予以删除。

### 2.1.6 算法结束条件

本算法以预先设定的迭代次数作为终止条件。

## 2.2 MFAIN 算法特点

对传统的人工免疫网络模型进行改进,改进的目的除了能对大样本、高维数的样本进行更好的特征提取之外,还能从存有噪声与误差的样本点中,提取原数据集有用的信息。因此,改进后的人工免疫网络算法具有很强的学习、识别、记忆和特征提取能力。

## 3 基于 MFAIN 算法的 RBF 网

### 3.1 RBF 神经网络

RBF 网络即为径向基函数网络,它是一种三层前向网络。第一层为数据输入层,第二层是隐层,第三层是输出层,如图 2 所示。

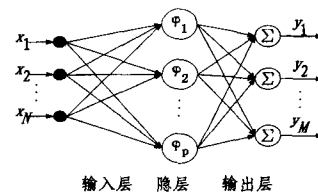


图 2 RBF 网络

图 2 中  $(x_1, x_2, \dots, x_N)$  为输入量,  $(y_1, y_2, \dots, y_M)$  为输出向量,输入层节点数、隐层节点数和输出节点数分别为  $N$ 、 $P$  和  $M$ ,隐层单元为高斯函数即

$$\varphi_i(x) = \exp(-\frac{\|x - c_i\|^2}{2\delta_i^2}) \quad (1 \leq i \leq P) \quad (9)$$

式中,  $c_i$  是隐层第  $i$  个单元的中心值,  $\delta_i$  为径向基函数的扩展常数。

RBF 网的输入层到隐层的权值一般都设定为 1,隐层到输出层的权值通过学习获得。RBF 网络的参数设计主要解决如何确定网络隐节点数,如何确定各径向基函数的中心值

以及扩展常数,以及如何获得输出权值。径向基函数的扩展常数是根据中心值的散布而确定的,为了避免每个径向基函数太尖或太平,本文将径向基函数的扩展常数设为

$$\delta_i = \frac{d_{max}}{\sqrt{2P}} \quad (10)$$

式中,  $d_{max}$  为各隐层中心之间的最大距离。对于隐层到输出层的权值  $W$  的更新,本文采用 LMS 算法。

由于 RBF 网的隐节点数对其泛化能力有极大地影响,因此寻找能确定隐节点数目的合理方法,是设计 RBF 网时需首先解决的问题。目前除聚类方法外,还有梯度训练方法、资源分配网络等一些方法,但这些都很难很好地提取代表输入样本集的特征点<sup>[6,10,11]</sup>,因此本文提出了将 MFAIN 算法与 RBF 网结合的方法,可以合理选取样本集的特征点作为 RBF 隐层的中心值,使得 RBF 网具有较好的泛化能力和较低的误差率。

### 3.2 基于 MFAIN 算法的 RBF 网的介绍

将输入样本点先通过人工免疫网络算法进行处理,获得的特征数据作为 RBF 网络隐层节点的中心值,然后利用 LMS 算法求出隐层到输出层的权值,其算法流程图如图 3 所示。

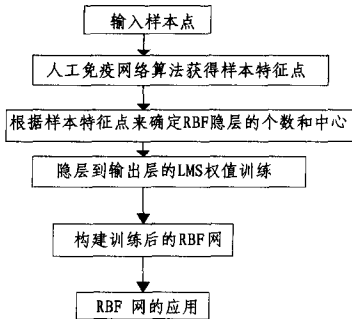


图 3 基于 MFAIN 的 RBF 网络流程图

## 4 实验仿真

为了验证 MFAIN 算法的特征数据提取功能,本文将基于 MFAIN 算法的 RBF 网用于函数逼近问题。先利用 MFAIN 算法提取输入数据集的特征数据,然后由 RBF 网实现函数逼近。通过比较使用 MFAIN 算法与不使用 MFAIN 算法的 RBF 网的实验效果,表明 MFAIN 算法的可行性与有效性。本实验是在 MATLAB2009b 环境下运行的仿真实验。

实验 1 函数为 Hermit 多项式

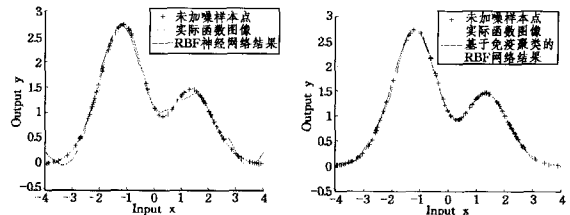
$$f(x) = 1.1(1-x+2x^2)e^{-\frac{x^2}{2}}, x \in [-4, 4] \quad (11)$$

算法初值设置如下: 抗体种群大小  $N=100$ , 进化代数  $gen=200$ , 抗体相似性系数  $\eta=0.8$ 。

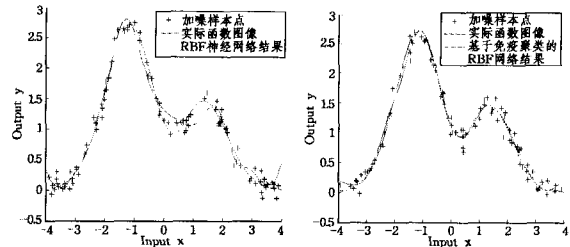
实验中,输入的样本点为 100 个,其实验测试结果的具体比较见表 1。在图 4 中,带“+”的点表示输入样本点,红色的实线表示函数的真实图像,蓝色的虚线表示逼近函数的图像。图 4(a)与图 4(b)中的输入样本均为未加噪样本,图 4(a)表示直接用 RBF 训练得到的逼近函数图像,4(b)表示用免疫网络模型算法和 RBF 结合得到的逼近函数图像。在图 4(c)和图 4(d)中,对输入的数据添加服从均值为 0,标准差为 0.1 的正态分布的噪声的干扰,图 4(c)表示直接用 RBF 训练得到的逼近函数图像,图 4(d)表示用免疫网络模型算法和 RBF 结合得到的逼近函数图像。

表 1 实验 1 结果的性能比较

逼近方法	输入样本数	是否加噪	隐节点数	测试数据数	平均误差
RBF 网络	100	否	11	100	0.0603
MFAIN 算法的 RBF 网	100	否	11	100	0.0442
RBF 网络	100	是	12	100	0.0987
MFAIN 算法的 RBF 网	100	是	12	100	0.0707



(a) 未加噪情况下 RBF 训练结果 (b) 未加噪情况下 MFAIN 算法的 RBF 网训练结果



(c) 加噪情况下 RBF 网训练结果 (d) 加噪情况下 MFAIN 算法的 RBF 网训练结果

图 4 实验 1 的结果

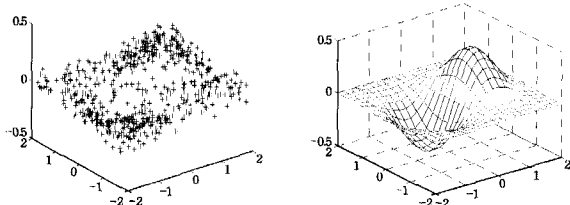
通过实验 1 可以得知,与一般的 RBF 网比较而言,基于 MFAIN 算法的 RBF 网具有泛化能力强、误差率低、速度快等特点,特别当输入数据集中包含有噪声和误差时,基于 MFAIN 算法的 RBF 网仍能较好地逼近函数,说明 MFAIN 算法具有平滑噪声、特征数据提取的功能。

实验 2 曲面为

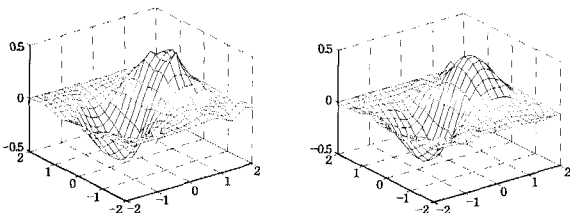
$$f(x, y) = xe^{-x^2-y^2}, x \in [-2, 2] \quad (12)$$

实验中种群的大小  $N=600$ , 进化代数  $gen=5000$ , 抗体相似性系数  $\eta=0.8$ 。

实验结果见图 5, 实验的性能比较见表 2。



(a) 加噪样本点 (b) 曲面的实际图像



(c) RBF 神经网络逼近曲面图 (d) MFAIN 算法的 RBF 网的函数逼近图

图 5 实验 2 的结果

CPU 占用现象	潜在问题
硬中断处理>系统调用	系统可能存在硬件问题,导致硬件中断频繁
上下文切换值过高	系统线程设计上可能存在问题,线程数过多导致上下文切换频繁
IOWAIT 过高,其他应用 CPU 利用率较低	存储系统可能存在问,导致 IO 延时过大,系统阻塞等待 IO 返回
CPU 平均队列深度>2	CPU 无法快速处理请求,导致请求堆积

根据 CPU 资源的使用情况判断是 IO 瓶颈还是 CPU 运算能力或数量不够,如 CPU 已无法快速处理请求,则需要更换更快的 CPU 或增加 CPU 的数量。

## (2) 内存

内存资源主要为 CPU 提供高速缓存,保证数据快速处理,分为物理内存和虚拟内存。

内存占用现象	潜在问题
内存使用 80% 以上(Linux 系统下重点关注 free page 指标)	内存资源被大量数据占用,即将成为瓶颈
虚拟内存占用>75%	内存资源基本耗尽,系统被迫采用磁盘空间作为缓存,内存资源成为瓶颈
页面请求延时过大	系统内存资源争用严重,导致申请页面请求出现等待
内存命中率低	应用系统设计不恰当或 OS 预取策略存在问题

查看 SWAP 区的利用率,若利用率高,表明 SWAP 区存在较频繁的页面换入换出,说明该内存偏少无法满足应用需求。若利用率低,说明内存容量可以满足。

## (3) 硬盘

硬盘使用现象	潜在问题
硬盘利用率>70%	硬盘资源已被大量占用,硬盘出现性能瓶颈
IO 延时过大	对于 OLTP,延时一般在 10ms 以内,不能超过 30ms,如延时过大,则说明硬盘已成为瓶颈
单盘 IO 队列深度>10	硬盘 IO 堆积严重,已成为瓶颈
业务实际下发 IO 请求数 > 存储侧接收 IO 请求数	IO 出现拆分,主机的卷管理设置或 HBA 卡设置可能出现问题

若检查出个别盘是慢盘,则更换该磁盘。

若每块物理盘都过载,说明磁盘阵列当前配置下的 IO 能

力无法满足应用对 IO 的需求,产生 IO 瓶颈,导致应用响应缓慢,处理时间过长。

若只有某个或 N 个逻辑卷片区存在与服务器 CPU 的“IO 等待”相对应的明显高的读写操作时段,说明该系统的 IO 请求超出了该卷的 IO 吞吐能力,使该磁盘组中的大部分磁盘过载,从而造成系统的服务器 CPU 进程时间都处于等待 IO 的闲置状态。这时需查看该逻辑卷片区上哪些数据文件被应用频繁访问,造成了访问过载,将访问过载的数据文件从该磁盘阵列上迁移至高性能的存储设备(SSD)上存储,或者分配更多的磁盘单元。

## (4) 通道

物理通道主要关注两个指标。

通道带宽利用率——利用率高达 90% 以上时,说明该传输通道已出现性能瓶颈,需扩大通道传输带宽或减少传输数据量。

通道每秒出现的数据传输错误次数——每秒出现 3 次以上的传输错误或误码时,说明传输通道受到干扰。

(5) san 网络模块较多,每层都可能是影响性能的短板,随着业务的不同,读写的 IO 大小和个数都不相同,那么每层设备的参数(如并发数)也会不同。若在某层出现 IO 堵塞情况,也有可能是该层的并发数设置较小,可适当增大。

**结束语** 本文分析了存储区域网络在联机事务处理下的性能研究,并提出了分析和解决性能瓶颈的方法,使读者可以快速准确地搭建一个完善的存储区域网络并对其进行性能调优,使整个系统能够高速稳定地运行。

## 参考文献

- [1] Barker R, Massiglia P. 存储区域网络精华-深入理解 SAN[M]. 舒继武,等译. 北京:电子工业出版社,2004:4
- [2] 熊军军. 架构应用软件性能测试问题研究[EB/OL]. <http://se.csai.cn/testnews/201006081348351467.htm>, 2010-06
- [3] 赵文辉,徐俊,周加林,等. 网络存储技术[M]. 北京:清华大学出版社,2005:14-20

(上接第 171 页)

在图 5 中,图 5(a)为加入幅度为 0.01 的白噪声的 600 个随机数据,即为加噪样本,图 5(b)为函数的真实图像,图 5(c)为 RBF 神经网络通过对加噪样本进行训练得到的函数曲面,图 5(d)为加噪样本经过免疫网络模型处理后再进行 RBF 神经网络训练得到的函数曲面图。

表 2 实验 2 结果的性能比较

逼近方法	输入样本数	是否加噪	测试数据数	平均误差
RBF 网络	600	是	600	0.0549
MFAIN 算法的 RBF 网	600	是	600	0.0404

通过实验 2 可以进一步说明,基于 MFAIN 算法的 RBF 网不仅对噪声有一定的平滑作用,而且存在误差率低、泛化能力强、速度快等特点。

**结束语** 针对目前人工免疫网络研究和应用的不足,本文给出了一种基于 MFAIN 算法的 RBF 网模型,在改进现有的人工免疫网络的基础上结合 RBF 网并加以应用。实验表明,MFAIN 算法具有减少数据冗余、平滑噪声、提取特征数据的功能,且应用中可以有效提高问题求解的速度和性能。

下一步工作将尝试将人工免疫网络模型算法与其他算法的性能进行比较,考虑与其他方法结合,并将其用于解决模式分类、系统辨识、函数逼近、模型预测等领域的一些工程问题。

## 参考文献

- [1] de Castro L N, Timmis J. Artificial immune systems as a novel soft computing paradigm[J]. *Soft Computing*, 2003, 7(8)
- [2] Timmis J, Knight T, De Castro L N. An overview of artificial immune systems[M]. *Natural Computation Sries*, Springer, 2004
- [3] Timmis J. Artificial immune systems—today and tomorrow[J]. *Sringer Science Business Media B. V*, 2006
- [4] 岳训,迟忠先,莫宏伟,等. 人工免疫网络模型的数据特征提取性能评价技术[J]. *小型微型计算机系统*, 2007(5)
- [5] Chen Yun-fang, Wang Ru-chun. A Classification Algorithm Based on Artificial Immune. *Chinese*[J]. *Journal of Electronics*, 2008
- [6] 谢景新,程春田,全磊光. 改进的免疫算法在函数优化中的应用[J]. *哈尔滨工业大学学报*, 2009(7)
- [7] 韩立群. *人工神经网络教程*[M]. 北京:北京邮电大学出版社, 2006
- [8] 耿生玲,康宝生. 基于 SOFM 与 RBF 神经网络的自由曲面重建[J]. *计算机工程与设计*, 2007(6)
- [9] 焦李成,杜海峰,等. *免疫优化计算、学习与识别*[M]. 北京:科学出版社,2007:17-170
- [10] 莫宏伟. *人工免疫系统原理与应用*[M]. 哈尔滨:哈尔滨工业大学出版社,2003
- [11] 阎平凡,张长水. *人工神经网络与模拟进化计算*[M]. 北京:清华大学出版社,2005