

基于智能互补策略的免疫算法

张利伟*, 苑津莎

(华北电力大学 电气与工程学院, 河北 保定 071003)

(*通信作者电子邮箱 zlwbd@126.com)

摘要:针对自组织抗体网络存在冗余抗体和网络性能不稳定的问题,提出一种基于智能互补策略的免疫算法。基于智能互补观点,该方法引入免疫进化算法中的免疫算子,它由接种疫苗和免疫选择两部分操作构成。接种疫苗利用 K -means 聚类算法抽取疫苗作为初始抗体,形成关于系统的粗略描述;免疫选择对记忆抗体进行优化,调整网络结构。在 Iris 数据集上的测试结果表明,该方法能够充分利用系统的先验知识快速有效地提取样本的数据特征,使得数据浓缩率和分类正确率更高。

关键词:人工免疫网络;模式识别;免疫算子;智能互补策略; K -means 聚类

中图分类号: TP183 **文献标志码:** A

Artificial immune algorithm based on intelligence complementary strategy

ZHANG Liwei*, YUAN Jinsha

(School of Electrical and Electronic Engineering, North China Electric Power University, Baoding Hebei 071003, China)

Abstract: There are redundant antibodies after training in self-organization Antibody Network (soAbNet) and its network performance is instable. In order to improve the performance of soAbNet, a hybrid immune diagnosis method was proposed based on intelligence complementary strategy. Immune operator was introduced into soAbNet, which consisted of two components: vaccination and immunoselection. Vaccines obtained through K -means algorithm were taken as initial antibodies in immune operator, and immune network architecture was optimized by immunoselection. The experimental results on Iris dataset demonstrate that, the proposed hybrid immune algorithm sufficiency makes use of prior knowledge and learns data characteristics effectively, and the diagnostic accuracy and data enrichment rate are higher compared with soAbNet.

Key words: artificial immune network; pattern recognition; immune operator; intelligent complementary strategy; K -means clustering

0 引言

人工免疫网络是模仿自然免疫系统功能的一种智能方法,它实现了一种受生物免疫系统启发,通过学习外界物质的自然防御机理的学习技术,提供了噪声忍受、无教师学习、自组织、记忆等进化学习机理,结合了分类器、神经网络和机器推理等系统的优点,从而提供了新颖的机器学习方法^[1]。其研究成果涉及控制^[2]、数据处理^[3]、优化学习^[4]和故障诊断^[5]等许多领域,已经成为继神经网络、模糊逻辑和进化计算后人工智能的又一研究热点^[6-8]。

自组织抗体网络是近年来提出的一种新型人工免疫网络,它忽略 B 细胞与抗体的区别,而将生物免疫系统抽象为顶点带权的不完全连通图^[9]。该算法无需人工设置任何参数与阈值,计算简便,易于实现,具有较强的适用性。目前自组织抗体网络已应用在模式识别和故障诊断领域,并取得了一定的成果^[10-12]。

但是通过分析可以发现,自组织抗体网络的学习算法中没有网络压缩机制,记忆抗体中存在很多冗余抗体,导致网络规模过大,网络效率降低;网络的初始抗体也是随机选取的,而学习过程和训练结果对初始抗体的选取比较敏感,容易造成网络性能不稳定,直接影响网络对抗原的识别能力。从现有文献和将来发展两方面看,充分利用问题的先验知识,采用

算法综合集成,都能有效提高人工免疫系统的实用性^[13]。

针对上述问题,本文提出一种混合免疫算法。该算法在自组织抗体网络的基础上引入了免疫进化算法中的免疫算子,免疫算子是由接种疫苗和免疫选择两部分操作构成的。基于智能融合思想为初始疫苗的获取提供了一条可行的途径,已有研究表明,结合其他智能策略(如神经网络和聚类算法),利用系统的先验知识构造疫苗,可提高免疫网络的效率,进而改善免疫算法的性能^[14]。本文利用 K -means 聚类算法对数据进行预处理,形成关于系统的粗略描述,以获得的疫苗作为网络的初始抗体。免疫选择对应于生物免疫系统中的免疫调节,实现对记忆单元的优化,构造相互协助又相互制约的网络结构,以保证网络的稳定。为了分析该算法的性能,本文从网络规模、接种疫苗对网络性能的影响两个方面在 Iris 数据集上进行了仿真实验。

1 自组织抗体网络^[7]

自组织抗体网络忽略抗体与 T 细胞的差异,将两者统一抽象为抗体,由若干不同类型抗体为节点构成的图,相同类型的抗体连接在一起,形成一个连通子图,每个抗体(节点)具有一定的浓度,浓度作为该节点的权值,表示了该抗体识别抗原能力的强弱,整个网络的结构可表示为一个带权的不完全连通图,如图 1。图中,有 3 类抗体,类别分别用 1, 2, 3 表示,

收稿日期:2012-10-16;修回日期:2012-11-15。 基金项目:国家自然科学基金资助项目(61204027)。

作者简介:张利伟(1983-),男,河北平山人,博士研究生,主要研究方向:智能信息处理、电气设备故障诊断;苑津莎(1957-),男,天津人,教授,博士生导师,主要研究方向:智能信息处理、电磁场。

同类抗体通过类别信息相连接, c 为抗体浓度。免疫抗体网络中每个抗体对各种抗原都具有识别、效应和记忆的功能, 抗体相互作用, 共同完成对抗原的学习、记忆和识别。

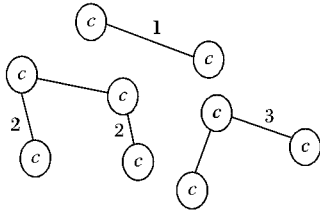


图1 自组织抗体网络示意图

1.1 抗原、抗体编码

免疫抗体网络中, 抗原、抗体编码方式采用实数编码, 抗原和抗体视为其形态空间中的点, 抗体与抗原的具有不同的编码结构, 赋予不同的含义。

1) 抗原编码。各类抗原视为其形态空间中的一个点, 用一个 n 维实数向量 Ag_j 来表示:

$$Ag_j = (Ag_{j1}, Ag_{j2}, \dots, Ag_{jn}) \quad (1)$$

2) 抗体编码。包括了抗体所属的类别、抗体的浓度和抗体属性信息三个部分。

抗体的结构可表示如下:

$$Ab_i = (T_i, C_i, Ab_{i1}, Ab_{i2}, \dots, Ab_{in}) \quad (2)$$

其中: T_i 是抗体 Ab_i 所属的类别, 即抗体 Ab_i 能够识别的抗原的类型; C_i 是抗体 Ab_i 的浓度, 表示抗体 Ab_i 能够识别的抗原的数量; $Ab_{i1} \sim Ab_{in}$ 分别表示抗体各维的模式特征。

1.2 三种学习与记忆策略

1) 抗体进化。网络中的抗体通过浓度和属性值的调整, 实现对抗原的学习和记忆。抗体 Ab_i 对输入抗原 Ag_j 的快速学习和记忆, 计算方法如式(3)所示:

$$\begin{cases} C_i' = C_i + 1 \\ Ab_{ik}' = (C_i \times Ab_{ik} + Ag_{jk}) / (C_i + 1); k = 1, 2, \dots, n \\ Ab_i = (T_i, C_i', Ab_{i1}', Ab_{i2}', \dots, Ab_{in}') \end{cases} \quad (3)$$

2) 抗体合并。网络中亲和度过高的两个同类抗体通过浓度和属性值的调整, 合并成一个新抗体, 压缩了网络规模, 降低了计算复杂性。具体地, 抗体 Ab_i 与 Ab_j 合并成为一个新抗体 Ab_k , 原有抗体 (Ab_i 和 Ab_j) 从自组织抗体网络中删除, Ab_k 按式(4)计算:

$$\begin{cases} C_k = C_i + C_j \\ Ab_{km} = C_i \times Ab_{im} + C_j \times Ab_{jm}; m = 1, 2, \dots, n \\ Ab_k = (T_i, C_k, Ab_{k1}, Ab_{k2}, \dots, Ab_{kn}) \end{cases} \quad (4)$$

3) 抗体新生。如果已有抗体无法正确识别输入抗原, 或已有抗体分布不合理, 网络自动生成一个新抗体完成对输入抗原的学习和记忆, 抗体浓度设为1。新抗体 Ab_i , 其取值有式(5), 其中 Ag_j 为输入抗原。

$$Ab_i = (T_j, 1, Ag_{j1}, Ag_{j2}, \dots, Ag_{jn}) \quad (5)$$

2 基于智能互补策略的免疫算法

通常, 人工免疫系统中初始抗体是随机产生的。自组织抗体网络中初始抗体也是由随机选取的训练样本构成的, 没有充分利用系统的先验信息。而网络的学习过程和训练结果对初始抗体的选取比较敏感。随机选择的初始抗体由于在形态空间中的分布不同, 造成训练中网络的收敛速度不同, 最终形成的记忆抗体质量也不相同。所以其网络性能不稳定, 对

抗原的识别能力有一定的影响。

自组织抗体网络存在一个较大的缺陷就是其抗体生成算法中没有网络压缩机制。随着训练次数的增加, 网络的规模将逐步增长, 运算速度将会降低, 训练后的网络中存在大量的冗余抗体。另外, 初始抗体数目的增加也会增大网络规模。但是在免疫系统中, 能够生成的抗体数目是有限的, 不可能无限地增长。

上述两个问题都限制了自组织抗体网络在模式识别中的成功应用。本文提出的免疫算法引入了免疫进化算法中的免疫算子。免疫算子是由接种疫苗和免疫选择两部分操作构成的。

接种疫苗是对待学习的训练数据进行具体分析, 按类别进行 K -means 聚类, 可获得有关数据的初始信息, 然后以聚类结果作为免疫疫苗, 直接植入免疫网络中作为初始抗体。各类别训练数据的聚类个数依据其复杂度确定。这里只是形成关于数据的粗略描述, 对聚类个数没有严格要求, 能反映数据分布特征即可, 为免疫系统提供先验知识。

免疫选择是在抗体生成算法中引入了免疫调节机制, 一般分两步完成。第一步是免疫检测, 即计算同类别抗体间的亲和度, 分析抗体的分布情况。第二步是抗体选择, 即亲和度值低于亲和度阈值的同类抗体发生抗体合并, 以简化网络结构, 控制网络规模。这里, 亲和度用欧几里得平方距离衡量, 距离越小, 亲和度越高。

本文免疫网络学习算法具体描述如下:

步骤1 接种疫苗, 并设定亲和度阈值 σ ;

步骤2 对于每一个递呈抗原 Ag_j , 计算其与网络中所有抗体的亲和度;

步骤2.1 选取识别抗体 Ab_i , 即与 Ag_j 亲和度最高的抗体;

步骤2.2 如果 Ab_i 和 Ag_j 的类别不一致, 网络发生抗体新生, 返回步骤2;

步骤2.3 Ab_i 和 Ag_j 的类别一致, Ab_i 即为最佳识别抗体 $Ab_b, Ab_b = Ab_i$;

步骤2.4 计算 Ab_b 与网络中同类抗体的亲和度, 选择与 Ab_b 亲和度最高的同类抗体, 记为 Ab_j ;

步骤2.5 如果 Ab_b 与 Ab_j 的亲和度小于 Ab_b 与 Ag_j 的亲和度, 则根据 Ag_j 发生抗体新生, 同时 Ab_b 和 Ab_j 发生抗体合并, 返回步骤2;

步骤2.6 如果 Ab_b 与 Ab_j 的亲和度大于 Ab_b 与 Ag_j 的亲和度, Ab_b 根据 Ag_j 发生抗体进化, 返回步骤2;

步骤3 免疫检测。计算同类别抗体间的亲和度, 亲和度值低于亲和度阈值 σ 的同类别抗体发生抗体合并;

步骤4 若满足终止条件, 停止训练; 否则, 返回步骤2。

上述算法中, σ 是亲和度阈值, 它通过调整网络中抗体的特异性水平来控制网络的可塑性和学习精度, 一般可取一个较小的值 (比如 $\sigma \leq 0.1$), 然后逐步增大其值, 并分析获得的结果, 以便最终确定一个合适的网络参数, 其值的选取以不降低分类正确率为前提; 终止条件与自组织抗体网络相同, 即连续的两轮训练, 网络中各类抗体的数目不再发生变化; 网络的输出是记忆抗体。

3 仿真实验

为了说明本文免疫算法在模式分类中的有效性, 仿真实验在国际 UCI 机器学习数据库中的 Iris 数据集上测试, 并与

自组织抗体网络进行了实验对比。

Iris(鸢尾属植物)数据集是用来进行分类问题测试的国际标准数据测试集,记录了三类鸢尾属植物(Setosa、Versicolor和Virginica)样本4个属性信息,每类植物50条,共计150条样本记录。其中,Setosa与其他两类鸢尾属植物差异较大,线性可分;而Versicolor和Virginica相似程度高,线性不可分。

在Iris数据集中,对每个类别的样本随机地选取40条,共120条样本,作为训练样本集;余下的30条样本作为测试样本集,其中每个类别的样本各10条。

模式分类的过程是:训练完成后,每一条测试样本即为输入抗原,计算其与网络中记忆抗体之间的亲和度,依据最邻近规则进行分类。

3.1 网络规模对比

为了测试抗体检测对网络规模的优化性能并便于和自组织抗体网络进行对比,这里将自组织抗体网络中选取的初始抗体作为疫苗。在自组织抗体网络中,各类别样本的初始抗体数为 k ,从每类训练样本中随机地抽取,初始抗体浓度设为1。本文免疫算法中,亲和度阈值 σ 取为0.6。对应初始抗体数 k 的5种不同取值,分别对两种网络进行测试,实验结果如表1和表2所示。

表1 不同初始抗体时自组织抗体网络的测试结果

k	训练次数	记忆抗体数目			分类正确率/%
		Setosa	Versicolor	Virginica	
4	10	4	15	17	100
8	4	8	10	14	100
12	2	12	16	16	100
16	2	16	19	18	100
20	2	20	20	22	100

表2 不同初始抗体时本文免疫算法的测试结果

k	训练次数	记忆抗体数目			分类正确率/%
		Setosa	Versicolor	Virginica	
4	2	4	6	8	100
8	2	4	6	11	100
12	2	4	7	10	100
16	2	4	8	13	100
20	2	3	6	12	100

从表1可以看到,随着初始抗体个数的增加,自组织抗体网络中各类别植物的记忆抗体数目也不断增加。与其他两类鸢尾属植物差异较大的Setosa,其记忆抗体数目由4上升到20,可见存在大量的冗余抗体。当 $k=20$ 时,网络的抗体总数上升到了62,比训练样本总数的一半还要多。其根本原因是由于自组织抗体网络中没有网络压缩机制,随着初始抗体个数的增加,网络的规模也将增大,训练后的网络中存在大量的冗余抗体。

表2是初始抗体数目不同时自组织抗体网络的测试结果。从表中可以看出,由于本文免疫算法中加入了免疫检测,记忆抗体数目受初始抗体个数影响不大,抗体总数保持在20左右,第一类植物Setosa的抗体数目更加稳定。对比两种网络的训练次数,本文免疫算法的收敛速度更快,只需经过2次训练即可完成抗体对抗原的学习和记忆;而自组织抗体网络在初始抗体数目为4时训练次数多达10次,大大降低了运算效率。

图2对不同初始抗体数目下两种免疫网络的记忆抗体总数进行了对比。通过比较可以明显看出,自组织抗体网络的

抗体数目随着初始抗体数目的增加呈线性增长,而本文免疫算法中的抗体数目基本保持稳定,远远小于前者。这说明本文提出的免疫算法通过免疫检测去除了网络中的冗余抗体,有效地压缩了网络规模,提高了网络的学习效率。

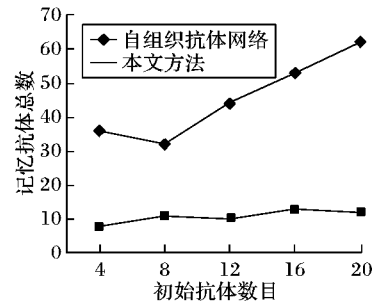


图2 不同初始抗体时的网络规模对比

为了更好地反映本文免疫算法在模式识别中的有效性,利用数据浓缩效果衡量指标进行算法分析,并与自组织抗体网络进行比较,比较结果见表3。数据浓缩效果衡量指标 E_n 为:

$$E_n = \left(1 - \frac{M}{N}\right) C \times 100\%$$

其中: M 为记忆抗体数目, N 为训练样本总数, C 为分类正确率。

表3 数据浓缩效果对比

k	数据浓缩效果对比		$\%$
	自组织抗体网络	本文方法	
4	70.0	85.0	16
8	73.3	82.5	55.8
12	63.3	82.5	48.3
			79.2
			82.5

通过表1~3可以看出不同初始抗体数目下,本文免疫算法保持和自组织抗体网络同样高的分类正确率,同时具有更高的数据浓缩率。

经上述分析可以得出结论,本文免疫算法中引入了免疫选择,通过网络压缩机制合理地调节了免疫网络的记忆抗体数目,优化了网络结构,同时有效地提取了抗原的模式特征,维持了免疫网络自身的稳定性。

3.2 接种疫苗对网络性能的影响

为了测试接种疫苗对网络性能的影响,将本文免疫算法按随机选取初始抗体和通过K-means聚类抽取疫苗两种方式分别在Iris数据集上进行了10次仿真实验并对比实验结果。随机选取初始抗体时,各类别样本的初始抗体数为 $k=4$,从每类训练样本中随机地抽取,初始抗体浓度设为1。抽取疫苗时,先对Iris数据集进行分析,Setosa与其他两类鸢尾属植物差异较大,线性可分;而Versicolor和Virginica相似程度高,线性不可分。把Setosa数据的聚类个数设为2,Versicolor和Virginica的聚类个数分别设为4,用K-means进行聚类得到10个聚类中心,作为网络的初始抗体,初始抗体浓度设为1。免疫检测中的亲和度阈值 $\sigma=0.6$ 。

图3~5给出了随机选取初始抗体时的网络实验结果。图3给出了10次实验的抗体数目变化情况,各类别植物的记忆抗体数目在小范围内产生波动。图4给出了10次实验中网络对测试样本集的分类正确率,对第一类植物Setosa和第3类植物Virginica的分类完全正确,对第二类植物Versicolor的测试样本有1例发生错误。图5给出了10次实验中网络对整个Iris数据集的分类正确率,可以看出对第一类植物Setosa的分类正确率很稳定,一直为100%,而对其他两类植

物的分类效果有所下降,这与数据集的数据分布一致;结合图4综合分析,可以知道网络对某些训练样本的分类发生了错误,没有完全提取训练样本的数据特征。

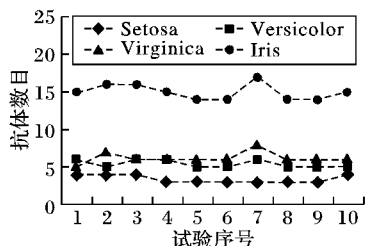


图3 随机选取初始抗体时记忆抗体数目情况

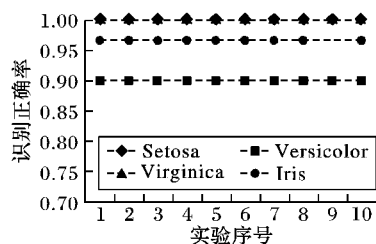


图4 随机选取初始抗体时对测试样本的分类正确率

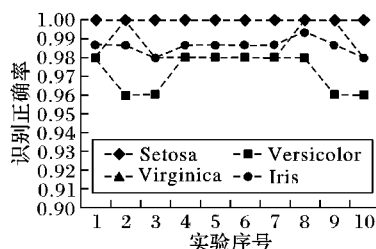


图5 随机选取初始抗体时对整体 Iris 数据集的分类正确率

通过聚类抽取疫苗在 Iris 数据集上进行的 10 次实验结果完全相同,如表 4 所示。每次实验经过 3 次训练即可完成对训练样本的学习与记忆,形成成熟的记忆抗体。利用得到的记忆抗体对测试样本集和整个数据集进行模式分类,分类正确率分别为 96.67% 和 99.33%,只对第二类植物 Versicolor 中的 1 条样本分类错误。

表4 抽取疫苗时的网络测试结果

植物类别	抗体数目	测试集识别率/%	整个数据集识别率/%
Setosa	2	100	100
Versicolor	5	90.00	98.00
Virginica	6	100	100
Iris	13	96.67	99.33

图6和图7分别对随机选取初始抗体和通过聚类抽取疫苗两种方式的实验结果进行了对比。

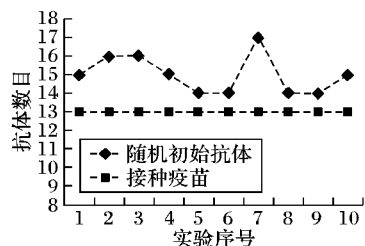


图6 10次实验中记忆抗体数目对比

从图6可以看出,抽取疫苗时网络形成的记忆抗体数目明显少于随机选取初始抗体时的记忆抗体数目,并且比后者更加稳定。图7对比了两种方式下网络对整个 Iris 数据集的分类正确率,抽取疫苗时网络的分类正确率明显高于随机选

取初始抗体时的分类正确率,并且比后者更加稳定。

从以上分析可以得出结论,通过对数据集的分析,利用对数据集的先验知识如类别和数据分布信息进行聚类,为网络接种疫苗,不但可以减小免疫网络的运算量,提高训练的收敛速度,而且使得网络性能更加稳定。

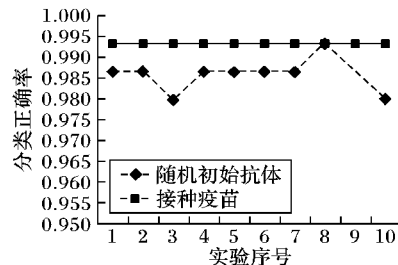


图7 10次实验中对 Iris 数据集的分类正确率对比

4 结语

本文提出的混合免疫网络引入了免疫算子,接种疫苗通过分析数据来提取疫苗作为初始抗体,充分利用了系统的先验知识,提高了网络的收敛速度,使得网络性能更加稳定;免疫选择通过网络压缩机制去除冗余抗体,有效地控制了网络规模,保证了算法的收敛性。实验结果证明,本文提出的免疫网络其算法效率和网络结构都优于自组织抗体网络,拥有更高的数据浓缩率和分类正确率。但是免疫选择中的亲和度阈值需要通过分析实验结果来确定,如何自适应地设定接种疫苗的数量和亲和度阈值的取值问题将是进一步研究的重要课题。

参考文献:

- [1] 丁永生, 任立红. 人工免疫系统: 理论与应用[J]. 模式识别与人工智能, 2000, 13(1): 52-59.
- [2] 胡珉, 吴耿锋. 基于免疫系统的多模型控制算法[J]. 计算机应用, 2008, 28(2): 297-301.
- [3] 潘章明. 基于亲和度累积的人工免疫网络聚类[J]. 计算机应用, 2011, 31(6): 1660-1663.
- [4] 张立, 晏琦. 基于高斯分布和模拟退火算法的免疫微粒群优化算法研究[J]. 计算机应用, 2008, 28(9): 2392-2394.
- [5] 侯胜利, 王威, 柏林, 等. 基于自组织免疫网络的传感器故障检测模型[J]. 计算机应用, 2009, 29(5): 1426-1429.
- [6] TIMMIS J, NEAL M, HUNT J. An artificial immune system for data analysis[J]. BioSystems, 2000, 55(1/2/3): 143-150.
- [7] de CASTRO L N, von ZUBEN F J. Immune and neural network models: theoretical and empirical comparisons [J]. International Journal of Computational Intelligence and Applications, 2001, 1(3): 239-257.
- [8] 焦李成, 杜海峰, 刘芳, 等. 免疫优化计算、学习与识别[M]. 北京: 科学出版社, 2003.
- [9] 李中, 苑津莎. 基于免疫抗体网络的模式识别方法: 中国, 101561880[P]. 2009-10-21.
- [10] 李中, 苑津莎, 张利伟. 基于自组织抗体网络的电力变压器故障诊断[J]. 电工技术学报, 2010, 25(10): 200-206.
- [11] 李中, 苑津莎. 基于免疫抗体生成算法的电力变压器故障诊断[J]. 华北电力大学学报: 自然科学版, 2009, 36(4): 25-29.
- [12] 李中. 基于人工免疫系统的电力变压器故障诊断技术研究[D]. 保定: 华北电力大学, 2010.
- [13] 焦李成, 杜海峰. 人工免疫系统进展与展望[J]. 电子学报, 2003, 31(10): 1540-1548.
- [14] 杜海峰, 王孙安. 基于 ART-人工免疫网络的数据浓缩方法研究[J]. 模式识别与人工智能, 2001, 14(4): 401-405.