

基于人工免疫网络的神经网络集成方法

张全平, 吴耿锋

(上海大学计算机工程与科学学院, 上海 200072)

摘要:提出基于人工免疫网络的神经网络集成方法 AINEN。在用 Bagging 生成神经网络集成之后, 将人工免疫网络的原理应用到神经网络集成, 组成了一个从微观上看是一个一个的神经网络, 而从宏观上看是一个大的人工免疫网络的集成。通过在微观层次上提高神经网络集成的个体之间的异构度, 在宏观层次上提高免疫网络的适应度, 从而降低集成的泛化误差。AINEN 与 GASEN 方法在标准数据集上进行的实验表明, AINEN 能取得更小的泛化误差。

关键词:人工免疫网络; 神经网络集成; 克隆; 变异

Neural Network Ensemble Method Based on Artificial Immune Network

ZHANG Quan-ping, WU Geng-feng

(College of Computer Engineering and Science, Shanghai University, Shanghai 200072)

【Abstract】 A method of Artificial Immune Network based neural network ENsemble(AINEN) is proposed. After initialing a network ensemble with Bagging, the artificial immune network is applied to the neural network ensemble which looks like an independent neural network in the microscopic view and an artificial immune network in the macroscopic view. Then the heterogeneity at microscopic level, and the fitness at macroscopic level is increased. As a result, the generalization error is decreased. The comparison of the results of AINEN and GASEN shows that the AINEN is more efficient than the GASEN.

【Key words】 artificial immune network; neural network ensemble; clone; mutation

1 概述

1.1 神经网络集成

上世纪 90 年代初, Hansen 等人^[1]在进行神经网络集成实验时, 发现了一个出人意料的现象, 那就是多个神经网络集成的泛化误差竟然比这些网络中最好的单个网络的泛化误差都要低。这个发现引来了众多专家学者的研究兴趣。在随后的 10 多年中涌现出了大量的对神经网络集成方法和应用的研究, 如神经网络继承融合^[2]。

神经网络集成一般分以下 2 步进行: 首先用训练数据生成一批单个的神经网络, 然后根据不同的任务确定获得输出的方法。Krogh 等人在 1995 年提出, 在进行神经网络集成时, 有 2 个因素决定着神经网络集成泛化误差, 一是网络个体的平均泛化误差; 二是网络个体之间的多样性。并给予了公式证明, 从此神经网络集成构造方法方面的研究工作集中于研究如何提高组成它的各网络个体之间的差异程度, 如 Bagging^[3], Boosting^[4]等。

针对如何提高神经网络集成的差异程度, 文献^[5]提出了选择性集成方法 GASEN, 为神经网络集成的发展开辟了新的天地。该方法的基本思想是, 如果能够为每个学习器指派一个权重, 而且该权重能够度量其对应的学习器在组成集成时的重要性, 那么就可以根据该权重来确定哪些学习器应该保留, 哪些应该被剔除, 例如, 权重小于某个预设阈值 λ 的学习器被剔除。

但 GASEN 是在已有神经网络的基础上剔除部分网络个体, 属于被动地提高集成的多样性。本文针对泛化误差和多

样性 2 个关键因素, 提出一种基于免疫网络的神经网络集成方法 AINEN, 首先用 Bagging 方法独立训练出一批神经网络, 然后将神经网络集成当作一个免疫网络, 利用人工免疫网络方法提高神经网络集成的个体的异构度和整个免疫网络的适应度。方法在执行过程中将通过免疫克隆, 主动地往集成中添加异构度和适应度高的神经网络个体来提高集成多样性。

1.2 自然免疫网络与人工免疫网络

1.2.1 免疫应答与免疫记忆

免疫应答(或免疫反应)是指抗原特异性淋巴细胞对抗原分子的识别、自身活化、增殖、分化及产生免疫效应的全过程。免疫应答又可分为初次应答和二次应答。

免疫记忆是指免疫系统能将入侵抗原反应部分抗体作为记忆细胞保留下来, 对于同类抗原的再次侵入时, 相应的记忆细胞被激活而产生大量的抗体, 缩短了免疫反应时间。

1.2.2 免疫网络理论

Jerne 最早提出免疫网络学说, 免疫网络学说的观点是建立在自身识别的基础上, 认为免疫系统淋巴细胞上分布的特异性抗原受体可变区(V)组成内网络, 通过免疫细胞相互识别 V 区的抗原决定簇来实现免疫系统的功能; 对外来抗原的应答, 是建立在识别自身抗原基础上的反应。

1.2.3 opt-aiNet

借鉴免疫网络理论提出的人工免疫网络可用式(1)描述,

作者简介:张全平(1981 -), 男, 硕士研究生, 主研方向: 机器学习; 吴耿锋, 教授、博士生导师

收稿日期: 2008-04-10 **E-mail:** qpingz@hotmail.com

整个人工免疫网络被看成是一个形式上稳定、内容上动态变化的网络。式中的“死去的未刺激细胞”通过只选择亲和度较好的细胞来实现，“复制的刺激细胞”通过抗原与抗体、抗体与抗体的共同作用来实现，且不区分 B 细胞和抗体。

$$\text{群体变化率} = \text{流入的新细胞} - \text{死去的无刺激细胞} + \text{复制的刺激细胞} \quad (1)$$

文献[6]根据免疫网络理论提出了一种名为 opt-aiNet 的人工免疫网络，该方法模拟免疫网络对抗原的刺激过程，主要包括抗体-抗原识别、免疫克隆增殖、亲和度成熟以及网络抑制。

2 基于人工免疫网络的神经网络集成 AINEN

在基于 opt-aiNet 人工免疫网络的神经网络集成方法 (Artificial Immune Network Based Neural Network Ensemble, AINEN) 中, 每个神经网络之间通过变异形成了一种既竞争又合作的关系。如图 1 所示, 每个椭圆表示一个神经网络个体, 椭圆内的数字表示该网络个体的名称和适应度, 每 2 个网络个体之间的连线表示这 2 个网络之间的异构度 (AINEN 中的适应度和异构度定义见下文)。图中网络个体 ANN1 与 ANN3 的异构度低于阈值, 且 ANN1 的适应度较低, 所以 ANN1 被剔除出集成; 同时将各自的输出, 或通过投票, 或通过加权平均最后得到集成的输出。这些网络个体之间相互竞争, 在微观层次上形成独立的输出域, 提高神经网络个体的异构度; 它们之间又相互合作, 在宏观层次上共同提高神经网络集成对目标输出的适应度。人工免疫网络中的神经网络个体受到 2 个因素的刺激: (1) 个体实际输出对目标输出的适应度; (2) 网络个体之间的异构度。只有那些适应度和异构度都高于指定阈值的网络个体才会被克隆, 而那些低于指定阈值的网络将被剔除出集成。图 2 描述了神经网络个体受到的刺激因素和做出反应的过程。

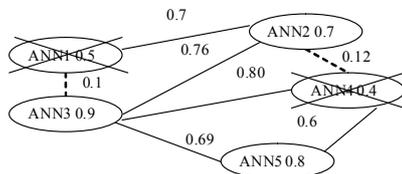


图 1 神经网络集成的免疫网络的竞争和合作示意图

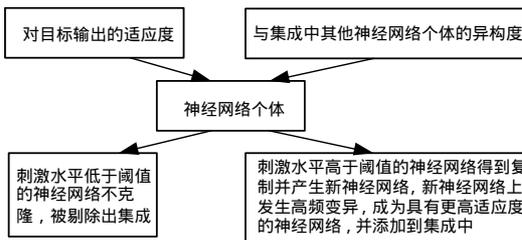


图 2 神经网络个体受到刺激的因素和做出的反应

在介绍 AINEN 方法之前, 先给出 AINEN 方法中的异构度 S 、适应度 B 和改进的变异算子 C 的定义:

(1) 异构度 S : 2 个独立的神经网络个体之间的差异程度。

与文献[6]不同, 本文介绍的 AINEN 将从集成中每 2 个神经网络个体的输出之间的关系来度量他们之间差异程度。

假设学习任务是利用 N 个神经网络对集成对 $f: R^m \rightarrow T$ 进行学习, 这里的 T 在分类任务中是类别标记的集合, 在回归任务中是某个随机分布, 或者某个函数的值域。令 f_i 表示第 i 个学习器的实际输出, $f_i = [f_{i1}, f_{i2}, \dots, f_{im}]^T$ 。其中, f_{ij} 表示第 i 个学习器在第 j 个示例上的实际输出。对于 2 个独立的

神经网络个体, 定义异构度矩阵

$$S = \begin{bmatrix} s_{11} & s_{12} & \dots & s_{1n} \\ s_{21} & s_{22} & \dots & s_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{n1} & s_{n2} & \dots & s_{nn} \end{bmatrix}, \quad s_{ij} = \text{dist}(f_i, f_j)$$

为 2 个神经网络个体 i, j 之间的异构度, 对于每个元素, 如果是分类任务, 那么使用海明距离计算而得; 如果是回归任务, 使用欧几里德距离计算而得。可以看出这是一个对称矩阵, 只需考虑上三角阵。

同时, 自然选择中的优胜劣汰的竞争关系在 AINEN 中主要表现为: 在从集成中剔除适应度低于阈值的网络个体的基础上, 对于异构度矩阵中的每个元素 s_{ij} , 如果小于指定的阈值 λ , 那么将 i, j 2 个网络个体中适应度较小的网络个体剔除出集成, 从而保留异构度和适应度高的网络个体。从分析多样性在神经网络集成中的作用可以看出, 提高多样性的目的就是当 2 个独立的神经网络对于同一个输入, 不能给出正确的输出时, 不要让它们都给出一样的错误输出。

(2) 适应度 B : 实际输出对期望输出的适应程度。

神经网络个体的适应度 B_i : 用 $A = [a_1, a_2, \dots, a_m]^T$ 表示期望输出, 其中, a_j 表示第 j 个用例的期望输出。

$$B_i = [b_{i1}, b_{i2}, \dots, b_{im}] = \frac{1}{\text{dist}(A, f_i)}$$

表示第 i 个神经网络个体的适应度。对于适应度高于阈值的网络个体将被克隆, 低于阈值的将被剔除出集成。

人工免疫网络的适应度 B_o : AINEN 中的人工免疫网络是一个集成神经网络, 集成神经网络的实际输出与期望输出存在着适应程度的问题, 也即 B_o 。AINEN 方法的目标就是提高免疫网络的适应度 B_o , 而 B_o 也是 AINEN 运行停止的标准。 B_o 的计算类似于计算异构度的方法。

(3) 变异算子 C : 用于对神经网络个体按照免疫原理进行变异操作。

在 AINEN 中, 定义遗传算子如下:

$$C = C + \alpha \times N(0, 1) \times \frac{e^{-\text{dist}}}{\lambda} \quad (2)$$

$$\alpha = \text{scale} \times (1 - \text{shrink} \times \text{currentGen} / n\text{Gen}) \quad (3)$$

其中, dist 是整个神经网络集成的输出与目标输出的距离; scale 用于控制变异的幅度; shrink 用于控制前一次变异与这次变异的变化幅度; currentGen 是当前变异到多少代; $n\text{Gen}$ 是变异的代数; $N(0, 1)$ 是高斯随机数生成函数。相对于原 opt-aiNet 方法中的变异算子, AINEN 的变异算子随着方法的迭代, 变化幅度将越来越小, 使方法更容易收敛。

AINEN 集成方法描述如下:

(1) 用 Bagging 方法生成初始的神经网络集成, 作为初始种群 $I\text{Abs}$ 。

(2) 计算每个网络个体实际输出与期望输出的适应度 B_i , 计算每 2 个网络个体之间的异构度 s_{ij} , 组成异构度矩阵 S 。

(3) 方法开始循环

1) 对集成中的神经网络个体进行克隆, 每个网络个体的具体的克隆数量由克隆数 cn_i 控制;

2) 对克隆出的神经网络进行实数编码;

3) 用变异算子对编码进行变异;

4) 将变异后的编码解码为新的神经网络;

5) 计算新网络个体的实际输出与期望输出的适应度 B_i ;

6) 从新网络中选出所有适应度高于父群体中最差的适应

度的个体加入到集成中；

7)计算每 2 个网络个体之间的异构度 s_{ij} ，组成异构度矩阵 S ；

8)对于集成中的任何 2 个网络个体 i 和 j ，如果他们的异构度 s_{ij} 小于阈值，则将其中适应度较小的网络个体剔除出集成；

9)计算整个集成的人工免疫网络适应度 B_o 。

(4)如果当前集成的适应度 B_o 没达到要求或者迭代次数没达到设定值，继续循环；否则，退出循环，方法结束，输出结果。

下面对 AINEN 方法作进一步说明：

在 AINEN 中将区分 B 细胞和抗体。这里 B 细胞对应于单个神经网络，抗体对应于单个神经网络的输出。为了减少不必要的克隆，每一次的迭代都会对集成中的网络个体计算异构度，只有异构度大于阈值的网络才被保留，这一点也与 opt-aiNet 不同。

克隆数(cn)是被选中的神经网络个体被克隆的数量。对于适应度不同的网络个体有不同的克隆数，计算公式如下：

$$cn_i = \text{ceil}(\log_n(b_i + \max(s_{i,j}))) \quad (4)$$

这样，克隆的数量会随着网络个体的适应度和异构度变化，异构度取该网络与其他网络异构度的最大值。同时，为了降低克隆数的敏感性，需要对适应度与异构度的和取对数。

为了对神经网络个体进行变异，需要对网络个体进行实数编码，具体的编码方式如下：

$$\underbrace{b_1, w_{11}, w_{12}, \dots, w_{1m}}_{\text{第1个隐节点的bias和权重}}, \underbrace{b_2, w_{21}, w_{22}, \dots, w_{2n}}_{\text{第2个隐节点的bias和权重}}, \dots, \underbrace{b_l, w_{l1}, w_{l2}, \dots, w_{lm}}_{\text{第l个隐节点的bias和权重}},$$

$$\underbrace{b_{l+1}, w_{l+1,1}, w_{l+1,2}, \dots, w_{l+1,l}}_{\text{输出节点的bias和权重}}$$

从方法中可以看出，在微观层次上从 2 个方面来提高集成的多样性：一方面剔除异构度低的网络个体，类似于 GASEN 方法；另一方面通过克隆、变异向集成中添加异构度和适应度高的网络个体，这属于主动地提高多样性。

3 实验

利用 Matlab 对 5 对取自 UCI 的数据做了分类实验，对 Freidman 和 Mexican 2 个函数做了回归实验，并分别与 GASEN 方法作了比较。

3.1 分类实验

首先独立训练出 10 个 BP 网络(每个网络有 1 个含 5 个隐节点的隐层)，然后在这个集成的基础上分别用 GASEN 方法和 AINEN 方法进行了比较。对于每个数据集，将数据按照 2:1 的比率分成训练数据和测试数据。神经网络的生成采用 GASEN 中的方法进行，即 2 个方法的实验都是基于同样的初始集成进行的。实验结果如表 1 所示。

表 1 GASEN 和 AINEN 方法在 UCI 数据集上的泛化误差比较

数据集	AINEN 泛化误差	GASEN 泛化误差	分类数
glass	0.051 3	0.076 9	7
wine	0.031 2	0.109 4	3
chess	0.073 7	0.076 3	2
soybean	0.544 3	0.582 3	19
wave	0.130 7	0.122 6	3

从实验结果可以看出，在 5 个分类数据集中，当分类数比较多时，AINEN 方法的泛化误差明显小于 GASEN 方法。这是由于 AINEN 既从宏观上考虑集成的适应度，又从微观上

考虑各个网络之间的异构度，当分类数比较多时，AINEN 能更深入地分析各个分类之间的差异，给出更精确的输出。

3.2 回归实验

对于回归任务，使用了 2 个比较常用的函数进行回归，一个是 Freidman#1 数据集，其数据由下式产生：

$$f = 10 \sin(\pi x_1 x_2) + 20(x_3 - 0.5)^2 + 10x_4 + 5x_5 + \varepsilon \quad (5)$$

另一个是 Mexican Hat 数据集，其数据由下式产生：

$$f = \frac{10 \sin(x)}{x} + \varepsilon \quad (6)$$

利用式(5)和式(6)分别随机产生 1 000 个数据作为训练数据，500 个数据作为测试数据。实验结果对比如表 2 所示。与分类任务不同，回归任务的实际输出可以是无穷多种，所以在进行计算时，由于 AINEN 有微观层次的计算，能考虑更多的细节，因此使得方法在 2 个回归任务上的泛化误差都比 GASEN 的低。

表 2 GASEN 与 AINEN 在回归数据集上的泛化误差比较

数据集	AINEN 泛化误差	GASEN 泛化误差
Freidman	0.000 05	0.000 1
Mexican	0.031 60	0.032 6

4 结束语

本文提出了一种基于人工免疫网络的神经网络集成方法 AINEN，并对该方法进行了理论分析，将该方法和 GASEN 方法做了对比实验。实验结果表明，AINEN 方法的效果比较好，尤其对于回归任务和分类数较多的分类任务，其获得的泛化误差要比 GASEN 小。

在实验过程中发现，由于集成中的个体通过变异以后，每个个体的适应度都有所提高，因此这时每个个体有更多正确的或接近正确的输出，或者说实际输出都更靠近目标，从而会使个体之间的相似度增加，也即个体之间的异构度降低。如果这时继续保持开始的剔除个体的异构度阈值，将造成网络个体过多地被剔除，从而造成泛化误差越来越大；而如果开始就把异构度的阈值定得太低，则会降低集成中网络个体的异构度，从而影响集成效果。此外，由于有更多的网络个体需要变异和计算，大大增加了计算量。因此，进一步的工作需要寻找一个合适的异构度阈值，或者让异构度阈值随着免疫方法的进行而自动调整。

参考文献

- [1] Hansen L K, Salamon P. Neural Networks Ensembles[J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 1990, 12(10): 993-1001.
- [2] 张晓丹, 赵海. 神经网络集成融合模型研究及应用[J]. 计算机工程, 2007, 33(14): 210-212.
- [3] Breiman L. Bagging Predictors[J]. Machine Learning, 1996, 24(2): 123-140.
- [4] Liu Yong, Yao Xin. Simultaneous Training of Negatively Correlated Neural Networks in an Ensemble[J]. IEEE Trans. on Systems, Man, and Cybernetics, 1999, 29(6): 716-725.
- [5] Zhou Zhihua, Wu Jianxin, Tang Wei. Ensembling Neural Networks: Many Could be Better than All[J]. Artificial Intelligence, 2002, 137(1/2): 239-263.
- [6] De Castro L N, Timmis J. An Artificial Immune Network for Multi-modal Function Optimization[C]//Proc. of IEEE Congress on Evolutionary Computation. [S. l.]: IEEE Press, 2002.