第 33 卷 第 8 期 计 算 机 工 程 2007 年 4 月
Vol.33 No.8 Computer Engineering April 2007
·人工智能及识别技术· 文章编号：1000—3428(2007)08—0201—04 文献标识码：A 中图分类号：TP393

基于特征加权的 Category ART 网络及应用

丁智国[1,2]，刘 悦[1]，吴耿锋[1]
(1. 上海大学计算机工程与科学学院，上海 200072；2. 浙江师范大学信息科学与工程学院，金华 321004)

摘 要：特征加权是特征选择的一般情况，它能更加细致地区分特征对结果影响的程度，往往能够获得比特征选择更好的或者至少相等的性能。该文采用自适应遗传算法来优化 Category ART 网络的特征权值，提出了一种改进的 Category ART 网络 FWART。在 UCI 标准数据集上的实验表明，FWART 网络获得了比 Category ART 网络更好的泛化能力。将该网络应用在地震震型预报上，取得了很好的预报效果。
关键词：Category ART 神经网络；特征加权；遗传算法；震型预报

Feature Weighted Based Category ART Network and Its Application

DING Zhiguo[1,2], LIU Yue[1], WU Gengfeng[1]
(1. School of Computer Engineering & Science, Shanghai University, Shanghai 200072;
2. School of Information Science & Engineering, Zhejiang Normal University, Jinhua 321004)

【Abstract】 Feature weighted is the general case of feature selection, which has better performance than (or at least has the same performance as) feature selection. In this paper, feature weighted is employed to improve the classification accuracy of the category ART networks. A novel network named FWART network is proposed, in which the self-adaptive genetic algorithm is used to optimize the weight vector. Experiments on the UCI datasets show that the FWART has better generalization power than Category ART. It is applied to predict the earthquake type. The result is satisfactory.
【Key words】 Category ART neural network; Feature weighted; Genetic algorithm; Earthquake type prediction

1 概述

特征加权(Feature Weighted)是特征选择的更一般情况。特征选择实质上就是用 0 或 1 对特征加权，采用这种方法导致该特征或参加或者干脆不参加结论的判断。特征加权能够更加细致地区分特征对结果影响的程度[1]，往往能够获得比特征选择更好的或者至少相等的性能。对属性加权，Punch 等人的结果证明用实数加权得到的错误率要比用离散值加权得到的错误率低；Komosinski和Krawiec等则提供了更进一步的证据证明特征加权性能优于特征选择[2]。现在将特征加权应用于提高学习器泛化能力的研究已经成为机器学习领域的一个研究热点。

本文将特征加权引入Category ART[3]网络，它是对传统的有监督模型Fuzzy ARTMAP的简化。通过一个模糊ART结构与分类算法，取代了传统模型中的双模糊ART结构，仍保留原ART网络的特性，适合于处理连续属性值的分类问题。该网络模型也存在一些缺陷，即假设样本各特征对输出结果的影响是相同的，在实际应用中，有些特征对预测结果可能有较强的决定性，而有些则发挥的作用较小。本文将特征加权引入Category ART(Adaptive Resonance Theory)网络，通过给每个特征赋予一个 0~1 之间的权重值来衡量其对结果影响的程度，从而提出一种基于特征加权的category ART神经网络 FWART(Feature Weighted based category Adaptive Resonance Theory networks)。如何获得特征的权值是首先需要解决的关键问题。文献[4]通过简单遗传算法SGA (Simple Genetic Algorithm)来确定最近邻分类器(Nearest Neighbor classifier)的特征权值，但SGA中交叉概率和变异概率的值是确定的。为了保证算法的稳定，变异率一般取值很小，因此算法实现复杂搜索时，经常出现不成熟收敛，即陷入局部极值点现象。文献[5]通过定义冲量基值来定性的判断每个特征的重要性状况，再使用简单遗传算法得到特征权值。本文采用自适应遗传算法来确定最佳的特征权值。自适应遗传算法中，交叉率和变异率可以根据适应度值的大小及个体和群体特征自适应地调整，保证了群体在整个进化过程中的多样性，从而提高了收敛速度和优化性能。

2 基于特征加权的 Category ART 网络算法

FWART 采用自适应遗传算法来获得特征权值，然后将这些特征权值赋给输入样本的每个属性，使得 Category ART 在进行模式节点选择匹配时，充分地考虑到了特征的重要性，从而使得网络的预测精度得到提高。FWART 算法具体描述如下：

输入 训练集 S，验证集 V，测试集 T，Category ART 算法。

输出 FWART 预测结果 R。

{初始化 FWART；
FW=1；//为每一个特征权值赋初值
$FWART_0$=Train(S, FW)； //按照初始权值训练Category ART
BestFW= AGA ($FWART_0$, V, FW)；//用自适应遗传算法(AGA)
//优化初始权值FW，得到最优权值BestFW；
R=Predict($FWART_0$, BestFW, T)；}

该算法的核心就是如何确定最优的特征权值。

基金项目：国家自然科学基金资助项目(20503015)
作者简介：丁智国(1979－)，男，硕士生，主研方向：机器学习，人工智能；刘 悦，博士、讲师；吴耿锋，教授、博导
收稿日期：2006-04-22 **E-mail：**ding_zhi_guo2005@163.com

2.1 Category ART 有监督网络学习

和其他所有的有监督 ART 网络一样，Categroy ART 网络包括识别、比较、查找和训练 4 个过程。

由于识别阶段主要是将输入向量I与每一个类模板Wj比较，并选择最大匹配的模板，因此如何选择最匹配模板成为决定Category ART性能的关键。本文对Categroy ART中的选择函数重新定义。新的选择函数$Choose_j$如下：

$$Choose_j(I)=\frac{|I\wedge W_j * FW|}{\alpha+|W_j|} \tag{1}$$

FW 为自适应遗传算法得到的特征权值向量。在网络初始学习阶段，对于属性 i，FW[i]=1。α 为选择参数，用于控制选择过程是更依赖于$|I\wedge W_j|$还是$|I\wedge W_j|/|W_j|$。$\wedge$为模糊运算符，定义为

$$(x\wedge y)_i=\min(x_i,y_i) \tag{2}$$

然后进入比较阶段，如果胜出节点 J 满足式(3)，则转入网络有监督学习阶段。

$$\frac{|I\wedge W_J|}{I}\geq\rho \tag{3}$$

有监督学习算法可用如下伪代码实现：

```
FOR all (Input I,TeachInfo T)     //输入样本 I，教师信息为 T
Learn (Input I, TeachInfo T)
IF T NOT in TeachInfo List
  Create TeachInfo T
  Add T to TeachInfo List
END IF
J = Categorize I by Categroy ART
IF LabelList[J] != T
  Temporarily increase ρ
  J=Categorize I by Categroy ART
  Reset ρ
END IF
  UpdateWeight(WJ)
  LabelList[J]= T
END Learn
END FOR
```

其中函数UpdateWeight(W_J)的调整为

$$W_J^{(new)}=\beta(I\wedge W_J^{(old)})+(1-\beta)W_J^{(old)} \tag{4}$$

式中，β称为学习率，β=1 称为快速学习。通常第 1 次更新时选取β=1，以后学习过程中取β<1。否则，会产生重置信号，将前一次胜出模板标记为无效，重新查找$Choose_j$最大的模板，直到匹配节点满足式(3)或者产生新的类模板。其中，ρ为警戒参数，用于控制模板个数，也就是分类的粗细程度。

2.2 自适应遗传算法优化特征权值

本文采用自适应遗传算法来优化特征权值，该算法的基本思想是根据适应度值的大小及个体特性和群体特征，自适应调整交叉率P_c和变异率P_m，使群体在整个进化过程中始终保持多样性，来达到改进遗传算法、提高收敛速度和性能的目的。

输入特征的权值 v 一般定义为 0 到 1 之间的实数($v\in[0,1]$)。0 表示该特征的重要性最小，1 表示该特征的重要性最大。在 FWART 中采用二进制进行编码，即将一个 0 到 1 之间的实数表示为一个 n 位的二进制字符串。本文的实验中采用 5 位二进制进行编码，即“ 00000 ”表示实数 0，“ 11111 ”表示实数 1.0。假设学习样本中输入特征的个数为 m，则在实验中先随机产生一个 5m 长的 0、1 串，每 5 位 0、1 串分别表示一个特征的权值，接着通过遗传算法的交叉和变异进化得到最优特征权值，最后通过式(5)将特征权值解码为 0 到 1 之间的实数作用于在学习阶段得到的 Category ART。式(5)中 l 是表示一个实数对应的二进制字符串的长度，g_i 为第 i 个基因。

$$v=\frac{\sum_{i=0}^{l-1}g_i 2^i}{2^l-1} \tag{5}$$

(1)交叉与变异

交叉率和变异率的选取对遗传算法的性能有重大影响。交叉率太小不利于基因重组，太大又容易破坏性能较优的个体。变异率太小有效基因缺失问题不容易得到解决，太大又易导致已有的有效基因丢失，当变异率达到 0.5 时，搜索便成为随机搜索。FWART 中根据适应度值的大小及个体特性和群体特征，自适应调整交叉率和变异率，使群体在整个进化过程中始终保持多样性，以达到改进遗传算法、提高收敛速度和性能的目的。为了能定量描述当前群体的分布特性，本文采用式(6)表示性能指标ϕ。

$$\phi=\frac{4\sum_{i=1}^{n}\sum_{j=i+1}^{n}H_{i,j}}{n^2 l} \tag{6}$$

其中，$H_{i,j}$表示编号为i和j的染色体之间的欧氏距离，n为群体大小，l为染色体长度。性能指标ϕ关于 0，1 百分比的函数图形如图 1 所示。

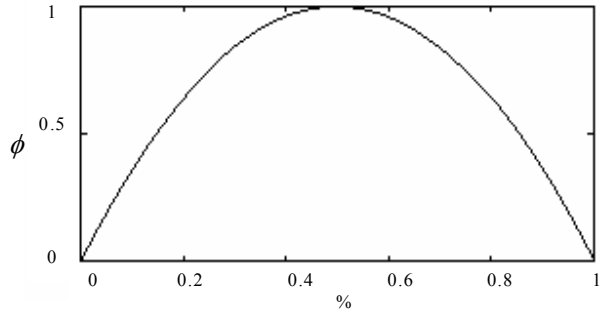

图 1 性能指标函数图形

由图 1 可知，$\phi\in[0,1]$：当$\phi=0$时，群体中的各个染色体完全相同；当$\phi=1$时，群体中各个染色体的每一位分别为 0 或 1 的可能性各为 50%，即是一个关于 0、1 的随机分布。

自适应调整交叉率和变异率的具体公式分别如式(7)、式(8)所示：

$$P_c=\begin{cases}\dfrac{f_{\max}-f'}{f_{\max}-f_{avg}} & f'>f_{avg}\\ 1.0 & f'\leq f_{avg}\end{cases} \tag{7}$$

$$P_m=\begin{cases}C(1-\phi+\dfrac{f_{\max}-f}{f_{\max}-f_{avg}}) & f>f_{avg}\\ 0.2 & f\leq f_{avg}\end{cases} \tag{8}$$

其中，$f_{\max}$为群体中的最大适应度，f_{avg}为群体平均适应度，f'为交叉的双方中适应度较大的个体的适应度，f为变异个体的适应度，C为适当的常数值，使P_m保持一定范围。

上述公式能保证在种群趋于最优时减小P_c和P_m的值，而在种群离散于解空间时增大P_c和P_m的值，适应度大的解有较小的交叉率和变异率，而适应度较小的解有较大的交叉率和变异率。保护较高适应度的解有利于算法收敛于全局最优解，而使适应度低的解有较高的交叉率和变异率，有利于防止算法“ 早熟 ”。因此自适应地改变交叉率和变异率，既避免算法收敛于局部最优，也防止优良的基因被破坏。

(2)计算适应度

适应度函数是衡量个体优劣的唯一标准，是遗传算法的关键步骤。本文所采取的策略是将种群中的个体反编码，得到特征权值，作用于网络的比较层中，这样每一个体 u 都会因不同的 FWARTu 而产生不同的预测结果，从而判断个体的优劣，也就是说对某个个体 u 将适应度函数定义为

Fitness(u)= FWARTu 的预测精度 (9)

3 UCI 标准数据集上的实验

3.1 实验设置

为了验证该方法的有效性，本文在 5 个 UCI 标准数据集(表 1)上比较了 Category ART，FSART(基于特征选择的 Category ART，即特征权值为 0 或 1)和 FWART 的性能。所有数据集都按大致 2∶1∶1 的比例分成了训练集、验证集和测试集 3 个部分。

表 1 数据集设置

数据集	样本总数	训练集样本数	验证集样本数	测试集样本数	输入属性数	分类数
Voting	231	131	50	50	16	2
Heart	270	140	65	65	13	2
Breast Cancer	683	343	170	170	9	2
Liver Disorders	345	175	85	85	6	2
Glass	214	114	50	50	9	6

3.2 实验结果及其分析

本文做了一组对比实验，实验重复了 50 次，所有结果均为 50 次的平均值，实验结果如表 2 所示。在数据集 Voting，Heart，Breast Cancer，Liver Disorders 和 Glass 上：3 个网络的外推精度为 FWART > FSART >Category ART。同时，50 次实验的标准差反映了 FWART 的稳定性好于或等于 Category ART。需要说明的是，不管是 FSART 还是 FWART 都以最少的模板赢得了学习算法在时间上的效率。Category ART、FSART 和 FWART 预报数据集预报精度的比较见表 2。

表 2 预报数据集预报精度比较(%)

数据集	Category ART		FSART		FWART	
	外推及标准差	模板数	外推及标准差	模板数	外推及标准差	模板数
Voting	92.80 ± 2.91	48	93.84 ± 4.02	38	94.04 ± 2.90	38
Heart	77.17 ± 4.98	51	77.78 ± 5.47	38	78.65 ± 4.95	38
Breast Cancer	96.47 ± 1.24	81	96.53 ± 1.1	56	96.60 ± 1.25	58
Liver Disorders	55.22 ± 5.21	28	56.41 ± 5.1	21	56.49 ± 4.59	21
Glass	52.40 ± 7.32	14	52.88 ± 7.54	11	55.24 ± 6.74	11

4 FWART 在地震震型预报中的应用

震后趋势判定是指破坏性地震或显著性地震发生后，对地震序列类型和后续强震或强余震进行判定或预报。一个较强地震发生后，在其附近可能发生一系列余震，一组在空间上和时间上丛集发生的地震称为一个地震序列。根据地震序列的类型，可以判断地震序列的发展趋势，从而对后续强震做出一定程度的预测。本文试图通过基于特征加权的 Category ART 网络方法来进行地震序列类型的判断。

根据不同的特征，地震序列的类型通常可分为 3 类：孤立型，主震-余震型和震群型。其中孤立型即在整个序列中有一个震级特别大的地震，称为主震，主震前的前震和其后的余震序列能量与主震相比微不足道，前震和余震数量少且震级相对小，我国地震序列中有 14%属于孤立型；震群型即在整个序列中有 2 个以上比较突出的大震，这几个大震震级相近，前震和余震一般比较多且震级相对大，震群型约占我国地震序列的 27%；主震-余震型即介于上述两种类型之间，有一个突出的主震，但其前震和余震序列比孤立型发育，但不如震群型，我国地震序列中主震-余震型的比例约为 59%。

地震序列类型的判别问题包括 2 个方面：(1)对一个完整序列的类型进行判定；(2)在序列发生后的早期对序列类型进行判定。本文实验前者进行了判定。

当前在序列类型的判断中，判断指标主要有序列最大地震与次大地震震级之差ΔM、主震能量E_m与整个地震序列能量E_Σ之比E_m/E_Σ(这两项指标具有明显的相关性[6])。根据地震研究的成果和专家经验，以下指标亦与地震震型预测相关。

(1)余震衰减系数 P

余震频次随时间 t 衰减公式为

$$N(t)=A/(t+c)^P \tag{10}$$

其中，N 为余震频次，A、c、P 为待定系数，其中 P 为衰减系数。在计算衰减系数 *P* 时，取时间间隔为 1 天(24h)。如果某一时间段内，地震次数为 0，则取以后十天内的总地震次数作为本段内的地震次数。

(2)b 值

b 值可以是整个序列的 b 值，也可以是余震 b 值。另外计算 b 值可利用线性最小二乘法和极大似然法。本节实验将充分考虑 b 值的 4 种不同计算方法，得到 4 组不同的数据，即

1)最小二乘法计算得到余震 b 值；

2)极大似然法计算得到余震 b 值；

3)最小二乘法计算得到的全序列 b 值；

4)极大似然法计算得到的全序列 b 值。

(3)序列归一化信息熵 k 值

序列归一化信息熵 k 值表示地震序列中地震能量分布均匀程度。

$$k=\frac{\ln S}{\ln N}-\frac{3.453}{\ln N}\frac{\sum \Delta_i S_i}{S} \tag{11}$$

其中，按震级大小将序列自大到小排列 $M_1 \ge M_2 \ge \ldots \ge M_n$。$\Delta_1 = M_i - M_n$，$i=1,2,\ldots,n$，$s_i = 10^{1.5\Delta_i}$，$S = S_1 + S_2 + \ldots + S_{n-1} + 1$。

(4)序列的主震震级 Mmax

根据文献[7]，序列 *b* 值与主震震级有关，其他参数与主震震级也可能有关。

(5)余震活动持续时间 *T*

余震活动持续时间表示地震序列中地震能量分布均匀程度。关于余震持续时间问题，不少学者做了大量的工作，周惠兰等曾确定余震序列中的起算震级 ML=2.0。根据余震频次衰减公式为

$$N(t)=A/t^P$$

定义ML=2.0 级地震衰减到每天发生 1 次时所相应的天数$T_{2.0}=10^{A/P}$为余震活动持续时间。对于一些余震序列，所记录的余震起始震级M_0可能大于ML=2.0 级，这时须按式(12)进行换算得到 2.0 级余震的持续时间。

$$T^*_{2.0}=10^{\frac{A-b(2.0-M_0)}{P}} \tag{12}$$

这里，A、P、b为根据余震起始震级M_0使用余震频次衰减公式和震级-频次关系得到的实际值。

$$N(t)=A/t^P$$

震级-频次关系为

$$LogN(M)=a-bM$$

根据前面的分析，本文取前述 7 项异常项目作为神经网络的输入属性，震群类型作为输出属性，共 3 类，即主震-余震型、震群型和孤立型。陆远忠、邓志辉、李胜乐等人研制的《基于 GIS 的新一代地震预报软件系统》中，共收集了 1966 年邢台地震以来发生在中国大陆地区的 5 级以上地震序列 183 次，形成地震序列目录索引，根据震级大小，该地震序列目录索引将地震序列命名为：

(1)G1~G22，分别代表 22 次 7 级以上地震序列；

(2)S1~S71 分别代表 71 次 6 级地震序列；

(3)M1~M90 分别代表 90 次 5 级地震序列。

按照该目录索引，计算得到 183 条学习样本，删除其中的 5 条异常样本，最终获得 178 条样本。根据 b 值的不同，设置如下 4 组试验，4 组试验数据均含有 178 条样本，按大致 2∶1∶1 的比例划分为训练集(98 条)、验证集(40 条)和测试集(40 条)。随机划分重复了 50 次，实验结果是 50 次的平均值。结果如表 3 所示。

表 3 Category ART 和 FWART 预报地震震型实验预测精度(%)比较

数据组	Category ART		FWART	
	内符	外推	内符	外推
第 1 组	86.68	81.38	90.64	85.48
第 2 组	86.94	82.60	90.30	84.25
第 3 组	86.59	79.95	90.10	84.25
第 4 组	87.17	82.70	90.17	85.80
4 组平均	86.85	81.66	90.30	84.94

由表 3 中的实验结果数据可知，在根据不同的 b 值设置的 4 组不同的试验中， FWART 在内符预测精度比 Category ART 预测精度分别提高了约 4%、3%、3%、3%；而在外推精度上也分别提高了约 4%、2%、5%、3%。在 4 组的平均内符精度提高了近 4%，在外推提高了 3%。总之，无论那组试验，FWART 都表现出了非常好的性能，4 组数据平均后，预测精度达到约 85%，得到了较满意的预测结果。

5 结束语

本文将特征加权引入 Category ART 网络(FWART)，通过对每个特征赋予 1 个 0~1 之间的权值来衡量其对结果影响的重要程度。本文引入自适应遗传算法优化传统的 Category ART 的权值，从而得到一个泛化能力更强的网络。在 UCI 数据集上的实验结果表明 FWART 比 Category ART 网络有更好的预测精度。在对地震震型预报的试验中，该方法也取得了令人满意的结果。进一步的工作包括对该方法的改进并融入神经网络集成的思想，使其能更好地应用于地震震型和震级的预报。

参考文献

1 Wen W X, Liu H. A Feature Weighting Method for Inductive Learning[C]//Proc. of the 3rd PRICAI. 1994: 338-344.

2 Komosinski M, Krawiec K. Evolutionary Weighting of Image Features for Diagnoses of CNS Tumors[J]. Artificial Intelligence in Medicine, 2000, 19(1): 25-38.

3 David Weenink, Categroy ART: A variation on Adaptive Resonance Theory Neural Net[C]//Proc. of IFA International Conference. 1997: 117.

4 Hussein F, Kharma N, Ward R. Genetic Algorithms for Feature Selection and Weighting, A Review and Study[C]//Proc. of ICDAR. 2001: 10-13.

5 刘 悦, 刘 辉, 李 远, 等. 基于冲量权值的 ART 神经网络及其在地震预报中的应用[J]. 计算机工程与应用, 2005, 41(5).

6 庄昆元, 王 炜, 黄冰树, 等. 地震序列类型的确定与现场预报规则的获取[J]. 中国地震, 2001, 21(3): 15-20.

7 焦远碧. 地震序列类型、地震序列 b 值与地震大形势关系初探[J]. 中国地震, 1998, 18(1): 33-40.

(上接第 200 页)

有唯一编号，可以直接查询知识库获得结果解释。

(2)对于神经网络推理，用“基于实例”的解释方法。可以选择一个学习范例作为当前求解实例的解释，选择的标准就是该学习范例与求解实例的某种逻辑距离最近。这相当于产生式系统中用一条模糊规则来解释推理得到的结论。

(3)通过网络分块技术实现解释。因采用了具有层次性的多个 BP 网络来解决规模较大的诊断问题，对于故障诊断系统针对每个典型故障建立一个 BP 网络，每个 BP 网络相当于一条大规则，其输入与规则的前提相对应，输出与规则的结论相对应，将 BP 网络所输入的对应征兆和输入值(可信度)展示给用户，用户即可知道故障所涉及的征兆，哪些被满足，可信度多少，结论的可信度是多少。

3 系统实际运行与结论

该系统于 2005 年 9 月试投运，在实际运行过程中，性能稳定，诊断准确可靠，采用专家诊断和神经网络推理相结合的诊断模式，既充分利用了专家的先验知识，又在此基础上发挥了神经网络的推理和对新输入模式的自学习能力，使得符号推理和数值计算推理并行进行，故障诊断更为准确、方便，减少了实际生产中的故障损失。本系统下一步工作是改进符号推理和数值计算推理结合机制，使二者相辅相成，以及对神经网络训练样本数据收集、处理的规范化，以提高网络输出精度和整个系统的诊断精度。

参考文献

1 鲍雨梅, 盛颂恩, 孙礼弘. 往复式压缩机故障诊断专家系统知识库设计[J]. 机械工程师, 2003, (2): 18-20.

2 魏少华, 陈效化, 隋巧梅, 等. 人工神经网络在车辆故障诊断中的应用[J]. 南京理工大学学报, 2005, 29(2): 193-196.

3 曹承志. 智能技术[M]. 北京: 清华大学出版社, 2004-09.

4 Ampazis N, Perantonis S J. Two Highly Efficient Second-order Algorithms for Training Feedforward Networks[J]. IEEE Transactions on Neural Networks, 2002, 13(5): 1064-1074.

5 肖元娇, 苏广川, 韩 雷. 基于神经网络的电子设备故障诊断专家系统[J]. 电光与控制, 2005, 12(3): 47-49.

6 Xu D, Wu M, An J W. Design of An Expert System Based on Neural Network Ensembles for Missile Fault Diagnosis[C]//Proceedings of the IEEE International Conference on Robotics, Intelligent Systems and Signal Processing, Changsha. 2003: 903-908.

7 王正群, 陈世福, 陈兆乾. 并行学习神经网络集成方法[J]. 计算机学报, 2005, 28(3): 402-408.

8 王金东, 张嘉钟, 刘树林. 应用神经网络识别往复式压缩机指示图[J]. 振动、测试与诊断, 2003, 23(3): 217-219.