

基于 ART1 改进算法的汉字分类研究

代小娟^{1,2} 林小竹¹ 周小正¹

(1 北京石油化工学院 自动化系, 北京 102617; 2 北京化工大学信息技术学院, 北京 100029)

摘 要 为了更好地利用自适应共振理论(ART1)来实现汉字的分类, 提出了一种改进的 ART1 算法。改进算法引用了同或的思想, 将输入模式与记忆模式相同的部分在输入模式总体中占的比例作为二者的匹配度, 从而降低了输入样本顺序对分类结果造成的误差, 增强了网络的稳定性和抗噪性。实验结果证明, 改进算法能够更好地将汉字进行分类, 求出聚类中心。

关 键 词 神经网络; 自适应共振理论; 警戒参数; 聚类中心

中图法分类号 TP183

自适应共振理论(adaptive resonance theory, ART)是美国波士顿大学学者 S. Grossberg 和 A. Carpenet 在研究人类认知过程中提出的一种无导师竞争型神经网络的学习机制。该理论的出现引发了系列实时无导师聚类学习算法的神经网络, 并且广泛应用于模式识别领域。经过不断发展和研究, 基本的 ART 网络模型包括: ART1、ART2 和 ART3^[1-3]。他们的共同特点是: 具有无监督学习功能, 能对时序信号进行实时学习、实时处理; 能对已学习过的对象快速响应、自动识别。这些特点使得 ART 神经网络非常适合于模式识别领域的应用, 因此它一出现就受到了广泛的关注。其中 ART1 用于处理二进制输入的信息; ART2 用于处理二进制和模拟信息这两种输入的信息; ART3 用于进行分级搜索。

ART 网络与常用的前馈型与反馈型的人工神经网络有所不同, 例如 BP 网络和 Hopfield 网络都不具备在线学习能力, 即这些网络的学习是分批进行的, 学习完毕后, 一旦有新的输入样本需要网络学习时, 必须把这些新的输入样本和以前的输入样本一起重新输入到网络中进行学习。而 ART 网络成功地解决了这一问题, 实现了网络的实时学习, 因而在模式识别的很多领域中得到了广泛的应用。

ART 网络是 ART 网络的第一代网络, 主要针对双极型或二进制输入模式的应用场合^[4-5], 其应用领域涉及故障诊断^[6]、模式分类^[7]、数据聚类^[8]、字符识别^[9]和人脸识别^[10]等。

在利用 ART1 网络进行汉字识别分类的过程中发现标准的 ART1 算法存在一定的缺陷, 当输入样本的输入顺序发生变化时, 网络的分类结果也随之改变, 即网络的学习不稳定, 对样本输入的顺序较为敏感; 同时对噪音比较敏感, 抗噪性不强。笔者针对这些问题提出了改进算法, 并在试验中通过改变警戒阈值对汉字实现了细分类和粗分类, 分类结果也比较理想, 并且在实现粗分类的同时提出了每类汉字的聚类中心。实验证明改进算法不但具有较强的稳定性和抗噪性, 同时粗分类的效果也比标准算法好。

1 ART1 人工神经网络

ART1 网络的基本结构如图 1 所示, 主要包括两个结构模块, 即注意子系统和取向子系统。取向子系统根据自下而上的无监督分类结果与外部环境匹配, 若“共振”发生, 则共振神经元权值得以调整; 反之, 取向子系统允许注意子系统增加其记忆容量以满足外部环境的需求。

由图 1 可以看出 ART1 网络是由比较层(F1)和识别层(F2)两层神经网络组成, 其学习

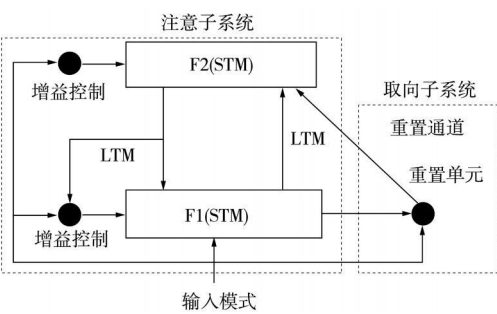


图 1 ART1 基本结构

及工作过程是通过反复地将输入学习模式由输入层向输出层自下而上地识别, 由输出层向输入层自上而下地比较来实现的。

当这种自下而上的识别和自上而下的比较达到共振, 即输入向量可以正确反映输入学习模式的分类, 且当网络原有的记忆没有受到不良影响时, 网络对一个输入学习模式的记忆和分类就算完成。将 ART1 网络用于分类问题时, 当出现一种新的类别时, 网络会通过在线学习, 将此类归入一种新的类别。

网络在训练过程中, 分类的粗细程度是由警戒参数决定的, 警戒参数设定得越大, 分类越精确。但是怎样设定警戒参数的大小, 目前还没有明确的标准, 一般在[0, 1] 之间取值。

1. 1 ART1 网络特点

ART1 神经网络的主要特点有:

- (1) 网络能实现实时在线学习, 具有自组织、自适应学习功能, 适应非平稳、非线性系统;
- (2) 对已经经过学习的对象具有稳定的快速识别能力, 同时对于新的对象能够迅速建立新的输出模式, 而且新增模式不受网络输出神经元的限制;
- (3) 具有自归一能力, 根据某些特征在全体中所占的比例, 有时作为关键特征, 有时又被当作噪声处理;
- (4) 当系统对环境做出错误反应时, 可通过提高系统的“警觉性”来迅速识别新的对象。

1. 2 标准 ART1 网络算法

标准 ART1 网络的算法叙述如下:

设输入层有 n 个神经元, 输出层有 m 个神经元, 输入样本为 $X_k = (x_1^k, x_2^k, \cdots, x_n^k)$, k 为样本个数, $W_j = \{w_{ij}\}$ 表示自下而上(由比较层到识别层)的前馈连接权值, $T_j = \{t_{ji}\}$ 表示从上而下(从识别层到比较层)的反馈连接权值, 其中

$i = 1, 2, \cdots, n, j = 1, 2, \cdots, m$ 。 ρ 为警戒阈值, 输入为二值。具体步骤如下:

步骤 1 初始化连接权(t_{ji} 和 w_{ij}) 以及警戒阈值 $\rho \in (0, 1)$ 。

$$t_{ji}(0) = 1 \tag{1}$$

$$w_{ij}(0) = \frac{L}{L - 1 + n} \quad (L \text{ 一般取 } 2) \tag{2}$$

步骤 2 将输入样本 X 输入到 ART1 网络的输入层。

步骤 3 计算识别层各神经元对应的加权和。

$$y_j = \sum_{i=1}^n w_{ij} x_i^k \tag{3}$$

步骤 4 选取获胜节点 J 。

$$Y_J = \max_{j=1}^m (y_j) \tag{4}$$

令获胜神经元 J 的输出 Y_J 为 1, 其余的为 0。

步骤 5 计算匹配度并进行警戒值检验。

$$|X_k| = \sum_{i=1}^n x_i^k \tag{5}$$

$$|T_J \circ X_k| = \sum_{i=1}^n t_{ji} \circ x_i^k \tag{6}$$

比较相似度:
$$\frac{|T_J X_k|}{|X_k|} \geq \rho \tag{7}$$

当式(7)成立, 则转入步骤 7, 否则转入步骤 6。

步骤 6 取消识别结果, 将输出层神经元 J 的输出值复位为 0, 并将此神经元排除于下次识别的范围, 返回步骤 4。当所有已存在的神经元都无法满足式(7)时, 选择一个新的输出神经元作为当前输入的分类结果, 进入步骤 7。

步骤 7 承认识别的结果, 调整连接权值。

$$t_{ji}(t+1) = t_{ji}(t) x_i^k \tag{8}$$

$$w_{ij}(t+1) = \frac{2 t_{ji}(t) x_i^k}{1 + \sum_{i=1}^n t_{ji}(t) x_i^k} \tag{9}$$

步骤 8 将步骤 6 复位的所有神经元重新加入到识别范围之内, 然后返回步骤 2 对下一个输入样本进行识别。

1. 3 ART1 网络算法的改进

一个样本模式是由样本中 0 和 1 分量的位置和数量共同决定的, 而 ART1 标准算法对匹配度的计算中, 只将相同 1 占输入模式中 1 的比例作为匹配度而忽略了 0 的匹配度, 使得网络对样本输入顺序和噪音相当敏感, 因而分类

和识别的结果都存在很大的误差。

余鹏飞等^[9]提出了一种将分量 1 和 0 的相似程度都考虑进去的改进算法: 将输入样本的维数 n 扩大为 $2n$, 从而构造了一个新的输入样本模式 $X'_k=[X_k,\overline{X_k}]$; 这样在计算匹配度时就把 1 和 0 的相似度都考虑了进去。由于该算法扩大了输入样本的维数, 从而增加了网络计算的复杂度, 针对这点提出一种新的改进算法。

改进算法中对匹配度的计算进行了重新定义, 引入了同或的概念, 使得改进算法更容易理解, 同时不增加输入维数, 减少了计算的复杂度。同或的逻辑定义为: 只有两个逻辑值相同时才为真。例如 X 与 T 的同或逻辑公式为: $T\odot X=(T\cap X)\cup(T\cap X)$ 。

由于 ART1 网络被用于二进制输入模式, 当比较匹配度时, 引入同或的思想, 将输入模式与记忆模式相同的部分在输入模式总体占的比例作为二者的匹配度, 从而把相同的 0 和 1 都考虑到匹配度的计算中, 避免了标准算法中的缺陷。

改进的 ART1 算法:

标准算法中其他步骤不变, 只将步骤 5(计算匹配度并进行警戒值检验) 作如下改变:

输入模式 X_k 和记忆模式 T_j 的同或为:

$|T_j\odot X_k|=|T_jX_k|+|\overline{T_j}\overline{X_k}|$ (10)

其中: $|T_jX_k|=\sum_{i=1}^n t_{ji}x_i^k$ (11)

$|\overline{T_j}\overline{X_k}|=\sum_{i=1}^n (1-t_{ji})(1-x_i^k)$ (12)

比较相似度: $|T_j\odot X_k|/n\geq\rho$ (13)

当式(13)成立, 则转入步骤 7, 否则转入步骤 6。

下面分别用标准算法和改进算法对汉字进行分类。

2 实验模拟研究

2.1 网络的输入样本

由于 ART1 网络适用于二进制输入模式, 所以实验采用的输入样本为标准汉字的 16×16 点阵形式, 即为 16 行 16 列的二维数组, 这里选取了 11 个具有代表性的三类汉字: 杠、机、枕、板、固、图、围、圈、汤、炆、杨。图 2 为汉字机、图和汤的点阵形式。

为了体现 ART1 网络的抗噪性, 这里在输

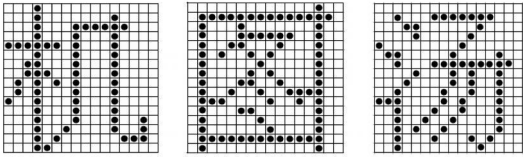


图 2 汉字的 16×16 的点阵形式

入样本中增加 3 个有噪音的汉字, 点阵形式如图 3 所示。

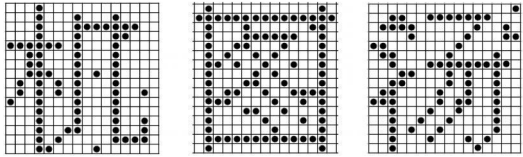


图 3 含噪音汉字的 16×16 的点阵形式

2.2 标准算法和改进算法的分类结果及分析
样本 1 的输入顺序为: 图、炆、围、枕、图(含噪音)、机、固、板、汤、杨、圈、机(含噪音)、杠、汤(含噪音)。

样本 2 的输入顺序为: 汤(含噪音)、杠、机(含噪音)、圈、杨、汤、板、固、机、图(含噪音)、枕、围、炆、图。即样本 2 为样本 1 的逆序。

表 1 和表 2 分别为标准算法和改进算法对不同顺序输入样本进行分类的结果。

表 1 ART1 标准算法对不同顺序汉字的分类结果

警戒参数 ρ	0.25	0.5	0.7	0.8	0.85	0.9	0.95	0.99
样本 1	3	5	8	9	12	13	14	14
样本 2	4	5	8	9	10	11	11	11

表 2 ART1 改进算法对不同顺序汉字的分类结果

警戒参数 ρ	0.25	0.5	0.7	0.8	0.85	0.9	0.95	0.99
样本 1	1	1	3	4	8	11	11	14
样本 2	1	1	3	5	8	11	11	14

2.2.1 输入顺序对分类结果的影响

观察表 1 可以看出, 对于不同顺序的输入样本, 标准算法的分类结果也不同。当警戒值 $\rho>0.9$ 时, 能区分样本 1 的 14 类汉字, 对样本 2 只能分为 11 类汉字。这是由于样本 2 中 3 个含噪音的汉字的顺序在标准汉字的前面, 当识别到标准汉字时, 根据标准算法计算的匹配度为 1, 所以, 无论 ρ 取多大, 当前标准汉字的匹配度都大于警戒值 ρ , 从而将标准汉字划分到属于含噪音的汉字的类别。由此可以得出输入样本顺序的改变会使得分类结果存在一定的误差。

表 2 中改进算法对不同顺序汉字的分类结

果则是一样的,从而减小了输入样本顺序对分类结果的影响。也就是改进算法对输入顺序不敏感,稳定性明显优于标准算法。

2.2 2 警戒值对分类结果的影响

由表 1 和表 2 可知,随着警戒值 ρ 的增大,网络对样本的分类越来越细。因此,当警戒值 ρ 取较大值时,是对输入样本进行细分类,反之则是对样本进行粗分类。

比较 ρ 取较大值两种算法对样本 1 进行细分类的结果可以看出:两种算法都能对样本进行分类,但抗噪性能不同。当 ρ 取 0.9 和 0.95 时标准算法将样本 1 分为 13 和 14 类,将含噪音字作为新的字,对分类和识别造成了一定的误差。而改进算法能准确地消除噪音汉字的干扰,将样本 1 分为 11 类,即改进算法的抗噪性能要好于标准算法。

2.2 3 不同算法对聚类中心的影响

两种算法对样本 1 进行粗分类时,由样本 1 的结构特征可以看出,较好的粗分类应分为 3 类。因此,这里比较样本 1 分 3 类时两种算法的分类结果。

- 标准算法 ρ 取 0.25 时分为 3 大类:
- 类别 1: 图、炆、围、图(含噪音)、固、汤、杨、汤(含噪音);
 - 类别 2: 枕、机、板、机(含噪音)、杠;
 - 类别 3: 圈。

由于粗分类时记忆模式 T_j 存储的是每个类别的共同特征,即为聚类中心。图 4 表示标准算法分 3 大类时各类的聚类中心。

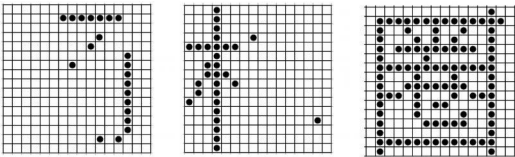


图 4 标准算法三大类的聚类中心

- 改进算法 ρ 取 0.7 时分为 3 大类:
- 类别 1: 图、围、图(含噪音)、固、圈;
 - 类别 2: 炆、汤、杨、汤(含噪音);
 - 类别 3: 枕、机、板、机(含噪音)、杠。
- 图 5 为改进算法分 3 大类时各类的聚类中心。

由两种算法的分类结果比较可以看出,标准算法虽然也分为 3 类,但是分类结果存在很大误差,主要体现在类别 1 和类别 3 没有区分

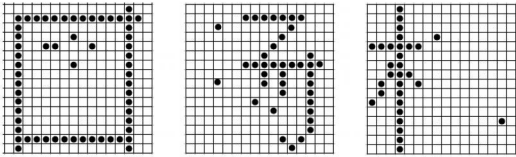


图 5 改进算法三大类的聚类中心

开;而改进算法准确地将输入样本分成 3 大类。比较两种算法的聚类中心可以看出,改进算法将每类汉字的共同部分提取了出来,从图 5 可以清楚看到 3 大类的结构特征;而标准算法因为分类的不准确导致了提取的聚类中心也存在一定的误差。

3 结论

改进的 ART1 算法考虑了 0 和 1 共同的匹配度,减小了输入样本顺序对分类误差的影响,使得网络的抗噪性大大增强,同时利用改进算法进行粗分类时还可以较为准确地提出每个类别的聚类中心,从而对 ART1 的分类原理有了更深一层的认识和理解。从分类结果可以看出改进的 ART1 算法能对汉字进行很好的分类,能够达到预期的分类效果。

参考文献

[1] CARPENTER G A, GROSSBERG S. The ART of adaptive pattern recognition by a Self-organizing neural network[J] . Trans IEEE on Computer, 1988, 21(3): 77-78.

[2] CARPENTER G A, GROSSBERG S. ART2: Self organization of stable category recognition codes for analog input patterns[J] . Applied Optics, 1987, 26(23): 4919-4930.

[3] CARPENTER G A, GROSSBERG S. ART3: hierarchical search chemical transmitters in self2organizing pattern recognition architecture [J] . Neural Network, 1990, 3(2): 129-152.

[4] 徐 勇, 荆 涛. 神经网络模式识别及其实现 [M] . 北京: 电子工业出版社, 1999: 257-309.

[5] 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: 220-241.

[6] 李 洪, 吴贻鼎. 利用 ART1 网络进行故障诊断的方法研究[J] . 电力系统及其自动化学报, 2003, 15(3): 23-27.

[7] 全朝阳, 石教英. 一种面向模式分类的修正的 ART1 神经网络[J] . 计算机学报, 1995, 18(9):

- 671-677.
- [8] 左 珑, 谭明峰. 基于 ART1 人工神经网络的数
据聚类[J]. 计算机工程与科学, 2002, 24(2):
92-95.
- [9] 余鹏飞, 刘 兵. 改进的 ART1 网络及其在英文
字符识别中的应用[J]. 云南大学学报, 2004, 26
(5): 397-400.
- [10] 周春光, 常 迪, 张 冰, 等. 一种改进的 ART1
算法及其在人像识别中的应用[J]. 小型微型计
算机系统, 1999, 20(10): 751-755.

Chinese Classification Based on Improved ART1 Algorithm

Dai Xiaojuan^{1,2} Lin Xiaozhu¹ Zhou Xiaozheng¹

(1 *Department of Automation, Beijing Institute of Petro-chemical
Technology, Beijing 102617; 2 School of information technology,
Beijing University of Chemical Technology, Beijing 100029*)

Abstract In order to make better use of adaptive resonance theory (ART1) to achieve the classification of Chinese characters, an improved algorithm of ART1 was proposed. The algorithm quotes the thought of XNOR, the percentage of the same part between the input and the memory mode in the overall input mode was used to define the match degree, thus the classification errors of the input sample sequence is reduced and the stability and the noise immunity of the network are enhanced. The experimental results proved that the improved algorithm can classify Chinese characters better and the cluster center can be calculated.

Key words artificial neural network; adaptive resonance theory; vigilance parameter; cluster center