

基于 ART2 改进算法的故障聚类研究

段霞霞,刘彦明,李小平,杨一展

DUAN Xia-xia, LIU Yan-ming, LI Xiao-ping, YANG Yi-zhan

西安电子科技大学 通信工程学院, 西安 710071

School of Communication Engineering, Xidian University, Xi'an 710071, China

E-mail: duanxiaxia@126.com

DUAN Xia-xia, LIU Yan-ming, LI Xiao-ping, et al. Study of fault cluster based on ART2 progressed algorithm. Computer Engineering and Applications, 2008, 44(15): 56-57.

Abstract: The Adaptive Resonance Theory 2(ART2), processing the analog signals of input patterns, is one of unsupervised clustering algorithms in neural networks. ART2 can reduce the pretreatment complexity of original data in data mining and improve the efficiency of mining. A progressed algorithm of ART2 which is composed of ART2 algorithm and K-Means algorithm can restrain cluster centers drift which is one of disadvantages of ART2 is developed. This method has been proved by simulation.

Key words: neural networks; cluster; Adaptive Resonance Theory2(ART2); K-means

摘要: ART2(自适应谐振理论2)算法是神经网络中一种可以对模拟输入信号或二值信号进行无监督聚类的算法,所以 ART2 算法能够降低数据挖掘中原始数据的预处理的复杂度,提高挖掘效率。针对 ART2 算法中出现的聚类中心偏移的缺点,采用 ART2 算法与 K-均值算法相结合的方法来抑制 ART2 中聚类中心偏移的现象。通过仿真对该方法进行了验证。

关键词: 神经网络; 聚类; ART2; K-均值算法

DOI: 10.3778/j.issn.1002-8331.2008.15.017 文章编号: 1002-8331(2008)15-0056-02 文献标识码: A 中图分类号: TP183

1 引言

神经网络具有能直接使用时间序列数据,处理高维数据,以及能够实时运行等特点,因此在智能故障诊断中应用很广。由于故障数据的形式往往是多种多样,这就需要在预处理阶段对数据做适当处理。但是如果对故障数据的预处理方法选择不当可能会使故障诊断算法诊断不出正确的结果。ART2(自适应谐振理论2)是 S.Grossberg 和 G.A.Carpenter 受到人类视觉系统工作原理的启发于 1987 年提出的可以处理模拟数据的神经网络聚类新算法^[1]。因为一些大型的电子机械系统中采集的数据往往是电流、电压、压强、速度等模拟值,所以 ART2 算法以其可以处理模拟数值的特点不但可以缩减预处理阶段所花费的时间,而且可以提高聚类的精确度。但是 ART2 算法由于其具有聚类中心偏移的缺点,所以其聚类功能受到了很大的限制。K-均值算法的前提是表示样本空间的聚类中的个数是预先知道的,这种假定本身限制了这一方法的利用。本文采取 ART2 与 K-均值算法相结合的方法对 ART2 聚类中心偏移起到了很好的限制效果,同时又克服了 K-均值算法聚类中心必须预先确定的缺点。

2 ART2 网络的基本原理

由图 1 可以看出 ART2 网络被分成两个大部分:取向子系统

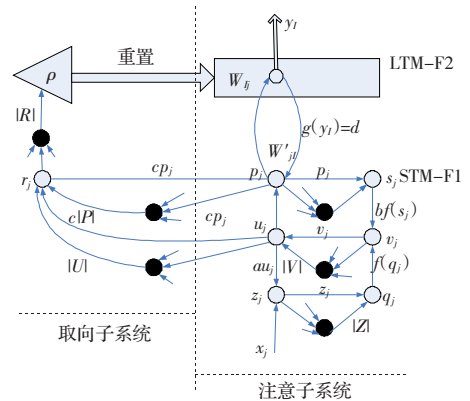


图 1 ART2 网络结构

统、注意子系统。其中注意子系统由 STM-F1 和 LTM-F2 层组成。

2.1 ART2 网络的基本原理

(1) STM-F1: 从图 1 可以看出 ART2 网络的 STM-F1 层由 3 个神经元层组成, STM-F1 采用 3 层结构的主要原因是消除输入数据中存在的噪声和加强网络中原有的存储模式与输入模式的对比。

(2) LTM-F2: LTM-F2 层的主要作用是选择一个输出神经

基金项目: 国家部委预研项目(the Pre-Research Foundation of China Ministries and Commissions)。

作者简介: 段霞霞(1981-),女,硕士研究生,主要从事数据挖掘,智能信息处理和故障诊断方面的研究;刘彦明(1966-),男,博士,教授,硕士生导师;李小平,女,硕士生导师,博士研究生,主要从事智能信号与信息处理,模糊神经网络,图像处理,新一代 IPV6 网络的研究;杨一展(1983-),男,研究生,从事数据库与数据挖掘,智能信息处理与故障诊断方面的研究。

收稿日期: 2007-08-29

修回日期: 2007-10-29

元存储的模式原型与输入模式相似的神经元作为输出节点。并把所选择的模式原型反馈到取向子系统以备重置子系统进行警戒测试。

(3)取向子系统:取向子系统的主要功能是完成模式对比和重置。如果 LIM-F2 输出节点所存储的模式原型与输入模式的相似度大于等于警戒值则调整 LTM-F2 层的当前激活节点所对应的向上权值向量和向下权值向量。否则,程序不继续向下执行并重新设置当前的激活节点,系统在剩下的输出节点中继续寻找最匹配的模式原型。如果输出节点都没有通过警戒测试,就创建一个新的输出节点(即在输出层增加一个新的神经元)来存储一个新的模式原型。

2.2 ART2 网络的权值学习规则与重置规则

(1)权值学习规则:ART2 网络的向上权值学习与向下权值学习规则分别为:

$$w_{ij} = \frac{u_j}{1-d}, w'_{ji} = \frac{u_j}{1-d} \quad (1)$$

由图 1 可以看出, u_j 是随着输入而改变的, 也就是说明每类的聚类中心是随着输入模式而改变的。

(2)重置规则:ART2 网络的重置规则为:

$$r_i = \frac{u_i + c p_i}{e + \|U\| + c \|P\|} \quad (2)$$

由此可以看出当向 ART2 网络输入一个样本时,网络通过比较它与网络中存储的模式原型的相似度来确定该样本属于那一类。

由式(1)、式(2)可以看出 ART2 网络的聚类中心会随着输入向量而产生偏移,并且本次输入的模式是在与上次输入模式比较的基础上划分类别的。当聚类中心偏移变大时,一些样本被划分到错误的类别中的可能性就变大。当聚类中偏移到一定程度时,会出现算法不能聚类的现象。

3 ART2 改进算法

首先假设 ART2 网络每个新的类的第一个输入为每一个新类的聚类中心 $\{S_1(l), S_2(l), \dots, S_k(l)\}$ 。其中 k 的取值是由每次聚类结果来确定的。

ART2 改进算法的基本步骤:

步骤 1 $z_j = x_j + au_j$

步骤 2 $q_j = \frac{z_j}{e + \|Z\|}$

步骤 3 $v_j = f(q_j) + bf(s_j)$

步骤 4 $u_j = \frac{v_j}{e + \|V\|}$

步骤 5 $p_j = u_j + \sum_{j=0}^{m-1} g(y_j)w'_{ji}$

步骤 6 $s_j = \frac{p_j}{e + \|P\|}$

其中:

步骤 7 $f(x) = \begin{cases} x, & x \geq \theta \\ 0, & x < \theta \end{cases}, g(y_j) = \begin{cases} d, & i=j \\ 0, & i \neq j \end{cases}$

步骤 8 $y_i = \max_j \{ \sum_{j=0}^{N-1} p_j w_{ij} \}$

步骤 9 $r_i = \frac{u_i + c p_i}{e + \|U\| + c \|P\|}, R = \|r\|$

如果当 $R \geq \rho - e$ 时(ρ 为预先设定的警戒值),则执行步骤

10,否则令 $y_i=0$,返回步骤 1。

步骤 10 计算 x_i 到聚类中心的距离:如果: $\|x_i - S_j(l)\| < \|x_i - S_i(l)\|$ 则 $x_i \in J$ 。

步骤 11 计算新的聚类中心: $S_j(l+1) = \frac{1}{N_j} \sum_{x_i \in J} x_i$

步骤 12 再把 $S_j(l+1)$ 作为 ART2 网络的新的输入执行步骤(1)~(9)。

$$w_{ij} = \frac{u_j}{1-d}, w'_{ji} = \frac{u_j}{1-d}$$

算法结束。

由 ART2 改进算法的基本步骤可以看出该算法从步骤 9 开始就于 ART2 算法不同了。ART2 改进算法与 ART2 算法的最大的不同点就在于 ART2 改进算法是通过比较输入模式与聚类中心的相似度的基础上划分输入模式的类别,而 ART2 网络通过比较网络上一次的输入模式与本次输入模式的相似度的基础上划分输入模式类别的。所以 ART2 改进算法能有效的克服聚类中心偏移的缺点,使聚类结果的正确率大大提高。

4 ART2 改进算法验证

某大型旋转设备常出现转子不平衡,转子不对中、油膜振荡、喘振和碰摩等故障^[2]。该设备的一个包含有 20 个样本的故障样本集^[2]的标准聚类中心如表 1 所示。

表 1 标准聚类中心

C1	C2	C3	C4	C5	C6	C7	聚类
0.077 2	0.059 9	0.102 5	0.811 4	0.045 7	0.151 2	0.027 0	不平衡
0.115 3	0.040 4	0.101 6	0.376 4	0.416 2	0.071 1	0.096 0	不对中
0.244 6	0.298 1	0.072 6	0.262 2	0.076 2	0.091 0	0.083 6	油膜振荡
0.142 9	0.282 0	0.050 7	0.352 8	0.098 5	0.101 5	0.231 3	喘振
0.116 1	0.083 2	0.076 6	0.548 0	0.076 6	0.225 4	0.103 7	碰摩

将表 1 中的设备标准故障样本集输入到参数 $a=10, b=10, c=0.2, d=0.83, e=0, \theta=0.1, \rho=0.986$ 的 ART2 网络通过仿真得到的聚类中心如表 2 所示。

表 2 标准故障样本集经 ART2 网络聚类得到的聚类中心

C1	C2	C3	C4	C5	C6	C7	聚类
0.090 8	0.048 6	0.154 0	0.702 8	0.121 6	0.116 3	0.060 3	不平衡
0.138 3	0.060 0	0.108 9	0.246 2	0.490 0	0.075 8	0.101 5	不对中
0.241 5	0.312 6	0.081 5	0.224 9	0.093 0	0.110 5	0.176 1	油膜振荡
0.114 2	0.252 6	0.028 6	0.445 2	0.077 7	0.072 4	0.121 5	喘振
0.115 6	0.112 7	0.105 9	0.414 1	0.100 9	0.354 4	0.133 6	碰摩

通过 ART2 网络对标准故障样本集进行聚类,结果样本 6,8,12,14,16,17,18 没有被聚类到正确的类别中,所以当 ART2 网络对该标准故障样本集的聚类的正确率为 65%,这样的聚类结果在故障诊断中是不符合要求的。可以看出与表 1 相比表 2 中不平衡故障的聚类中心发生了严重的偏差,这是导致样本 6,8,17,18 被聚到错误的类别(不平衡)的最基本的原因。该标准故障样本集的样本 1,2,3,4,6,8,的相似程度是逐渐逼近样本 17 的,因为 ART2 网络中本次输入的模式是以上次输入模式的比较的基础上划分类别的,所以聚类中心也是一步一步向样本 17 逼近的。由此可以得出 ART2 网络聚类中心偏移的缺点导致样本 17 被错误的划分类别。