

基于交叉熵和新转移函数的模糊神经网络分类器*

毛玲 孙即祥 季虎

(国防科技大学电子科学与工程学院 湖南长沙 410073)

摘要 针对目前普遍采用的误差平方和准则及 Sigmoid 转移函数在 BP 算法应用中存在的缺陷和不足,提出了基于交叉熵准则和新的 S 型转移函数构建的模糊神经网络分类器,并将这种分类器应用于心肌梗死的定位诊断,结果表明其训练效率和识别性能都明显优于传统的模糊神经网络。

关键词 交叉熵;转移函数;模糊神经网络分类器;心肌梗死

中图分类号:TP183 文献标识码:A

Fuzzy Neural Network Classifier Based on the Cross Entropy Rule and the New Transfer Function

MAO Ling, SUN Ji-xiang, JI Hu

(College of Electronic Science and Engineering, National Univ. of Defense Technology, Changsha 410073, China)

Abstract In view of the fact that BP algorithm based on the common used error square sum rule and Sigmoid transfer function have some limitation and shortcomings, the cross entropy rule and a new transfer function are adopted for constructing and training process of the fuzzy neural network classifier. It is used to realize the orientation of myocardial infarction, and the results prove that this classifier has the capability of outperforming the traditional fuzzy neural network in training efficiency and recognizing ability obviously.

Key words cross entropy; transfer function; fuzzy neural network classifier; myocardial infarction

基于误差平方和准则及 Sigmoid 转移函数的 BP 神经网络已被广泛使用,并成功地应用于模式的分类识别等诸多方面。BP 算法是通过使目标函数最小来实现网络优化的一种算法。从优化的观点看,目标函数是至关重要的,在 BP 算法中用误差平方和作为目标函数对分类问题并不是最合理的^[1],误差平方和准则是在应有输出满足高斯分布的条件下由最大似然原则导出的,而在分类问题中,输出并不满足高斯分布,而且传统 BP 算法收敛速度慢、产生局部极小点的缺陷也与采用误差平方和准则有关。另外, Sigmoid 函数取不到 0、1 值,同时导数值范围很窄造成了网络学习速度慢^[2],给网络在模式分类方面的应用带来不足。本文提出选用更适用于模式分类的交叉熵准则和新的 S 型转移函数,应用于模糊神经网络分类器的构造与训练。

1 模糊神经网络分类器结构

所有将模糊逻辑与人工神经网络结合起来的混合智能处理系统,都可以称作是模糊神经网络。而实际上,采用不同结合方式的模糊神经网络之间有着很大的不同,本文采用了结构等价型结合方式,即用等价的神经网络来表示模糊逻辑系统。这样,既可以利用模糊推理技术指导神经网络结构的建立和权系数的初始化,又可以利用神经网络实现模糊系统性能的进一步完善。

在模糊系统中,“模糊化”→“模糊推理”→“模糊判决”是构成模糊系统的最基本模块。将模糊系统表达成联接主义方式的网络结构,就得到模糊神经网络。模糊神经网络分类器用多层前馈网络实现,如图 1 所示。结构模型可以概括为三部分:

(1) 模糊预处理部分。包括了输入层和模糊化层。输入层节点的个数等于输入样本的特征维数。

* 收稿日期:2004-06-19
作者简介:毛玲(1980-),女,硕士生。

输入的各特征在模糊化层被分别映射到相应的几个模糊集,得到其相应的隶属度^[3]。其中,每个特征所对应的模糊集的个数可根据特征本身的属性及推理的需要来确定,不同的特征对应的模糊集的个数可以不同。隶属函数选用常见的带参数的分布函数近似表示,所选函数的形态应尽量符合模糊变量的本质特征,其中的参数可在网络学习中逐步修正改进。

(2)模糊推理部分。神经网络分类器的核心部分,联系着模糊推理的前提与结论,具体实现样本的分类识别。该部分每个神经元的输出值都代表着某条推理规则前提或结论的模糊状态,有着明显的物理意义。模糊推理部分包含的网络层数没有限定,一切根据推理的具体情况而定。

(3)输出部分。输出部分只有一层,层内节点的个数等于模式类别数。因为希望每个输出代表样本属于某个类别的概率,该部分可用来实现由推理所得结论的模糊状态值到概率的转化。

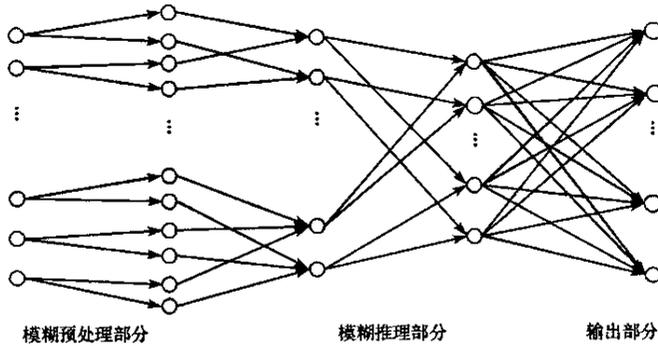


图1 模糊神经网络分类器的结构示意图

Fig.1 The structure of fuzzy neural network classifier

2 交叉熵函数准则及训练算法

最大似然方法是最估计方法之一,下面就由最大似然原则导出交叉熵函数准则。

对于一组独立同分布的训练样本 $\{x(n), t(n)\}^N$,其中 x 为输入数据, t 为相应的目标值, n 表示样本序号, N 表示样本数,其似然函数为

$$L = \prod_{n=1}^N p(x(n), t(n)) = \prod_{n=1}^N p(t(n) | x(n)) p(x(n)) \tag{1}$$

使似然函数 L 最大等价于使 L 的负对数最小,也就是使(2)式最小:

$$E = -\ln L = -\sum_n \ln p(t(n) | x(n)) - \sum_n \ln p(x(n)) \tag{2}$$

由于第二项 $(p(x))$ 与网络参数无关,故在求最小时可省去。这样就得到

$$E = -\sum_n \ln p(t(n) | x(n)) \tag{3}$$

设用作分类器的神经网络有 c 个输出,每个输出表示样本属于相应某一类的概率。令 $t_k(\cdot)$ 表示第 k 个神经元的期望输出,当 $x(n)$ 属于第 l 类(用 ω_l 表示)时, $t_k(n) = \delta_{kl}$ 。给定输入 $x(n)$,第 k 个神经元的实际输出即是 $y_k(n) = P(\omega_k | x(n))$,所以有

$$P(t(n) | x(n)) = \prod_{k=1}^c (y_k(n))^{t_k(n)} \tag{4}$$

将(4)式代入(3)式,即得

$$E = -\sum_n \sum_{k=1}^c t_k(n) \ln y_k(n) \tag{5}$$

这就是输出 $y(x)$ 与目标 t 的分布间的交叉熵,当且仅当 $y_k(n) = t_k(n)$ 时(5)式取最小值。

可以证明,基于交叉熵准则的BP算法与基于误差平方和准则的BP算法相比,对微小变化的反应

灵敏度高,收敛速度更快,不易产生振荡,且易跳出局部极小点^[4];并且,交叉熵函数是从模式分类的角度推导出来的,因而更适于作为网络分类器的目标函数。

交叉熵准则要求网络的输出表示样本属于各类的概率,所以需要固定所有输出值之和为1。设推理部分最后一层第 k' 个节点的输出为 $y_{k'}$, 则输出层第 k 个节点的输出为

$$out_k = \frac{\exp(y_k)}{\sum_{k'} \exp(y_{k'})} \quad (6)$$

上式是一种归一化了的指数函数,可称为 Softmax 函数^[1]。

对于一个特定的样本,其交叉熵为

$$E(n) = - \sum_{k=1}^c t_k(n) \ln out_k(n) \quad (7)$$

可以得到

$$\frac{\partial E(n)}{\partial y_k} = out_k(n) - t_k(n) \quad (8)$$

下面给出网络模糊推理部分在权值训练中用到的基本计算公式^[5]:

(1) 前向计算

对第 l 层的第 j 个单元

$$v_j^l(n) = \sum_i w_{ij}^l(n) y_i^{l-1}(n) \quad (9)$$

$$y_j^l(n) = f(v_j^l(n)) \quad (10)$$

(2) 反向计算

对最后一层的第 k 个单元

$$\delta_k^l = \frac{\partial E(n)}{\partial v_k^l(n)} = \frac{\partial E(n)}{\partial y_k^l(n)} \frac{\partial y_k^l(n)}{\partial v_k^l(n)} = [out_k(n) - t_k(n)] f'(v_k^l(n)) \quad (11)$$

对中间层的第 k 个单元

$$\delta_k^l = \frac{\partial E(n)}{\partial v_k^l(n)} = \frac{\partial E(n)}{\partial y_k^l(n)} \frac{\partial y_k^l(n)}{\partial v_k^l(n)} = \left[\sum_j \delta_j^{l+1}(n) w_{jk}^{l+1}(n) \right] f'(v_k^l(n)) \quad (12)$$

权值修正量

$$\Delta w_{ij}^l(n) = -\eta \frac{\partial E(n)}{\partial w_{ij}^l(n)} = -\eta \frac{\partial E(n)}{\partial v_k^l(n)} \frac{\partial v_k^l(n)}{\partial w_{ij}^l(n)} = -\eta \delta_k^l y_i^{l-1}(n) \quad (13)$$

对模糊化层隶属函数参数的调整,也采用基于梯度下降的反传算法,具体的计算公式由隶属函数所采用的函数形式决定。

为了加速学习,训练中还可采用增加动量项及可变步长、改变训练样本的输入次序等方法。

3 转移函数

在保证网络训练收敛的前提下,权值调整的幅度越大,学习速度就会越快,而由权值训练所用到的基本公式可看出,权值调整的幅度与转移函数的导数成正比。该导数在网络学习过程中被大量地使用,会对整个网络的学习过程产生很大的影响。Sigmoid 函数具有良好的平滑渐进性,是目前在 BP 网络中最常使用的一种转移函数。但是,它的导数值范围很窄,限定在 $0 \sim 0.25$,较小的导数值导致了网络学习速度较慢。而且,它的函数值介于 $0, 1$ 之间,达不到极限的 0 或 1 ,因而无法准确表达模式分类中已确定类别的样本的类属概率。

针对 Sigmoid 函数的缺点和不足,本文提出一种新的 S 型转移函数,表达式如下:

$$f(x) = \begin{cases} 1, & x > \frac{\pi}{2} \\ \frac{\sin x + 1}{2}, & -\frac{\pi}{2} \leq x \leq \frac{\pi}{2} \\ 0, & x < -\frac{\pi}{2} \end{cases} \quad (14)$$

它的导数表达式为

$$f'(x) = \begin{cases} \frac{\cos x}{2}, & -\frac{\pi}{2} < x < \frac{\pi}{2} \\ 0, & \text{其它} \end{cases} \quad (15)$$

此函数满足反传算法对转移函数可微的要求,且具有比 Sigmoid 函数更陡峭的函数曲线,导数值范围扩大到 $[0, 0.5]$,可以获得较快的网络学习速度;其值域为 $[0, 1]$,可以取到 0 和 1。新的 S 型转移函数(用 Newstf 表示)与 Sigmoid 函数的函数曲线比较如图 2 所示。

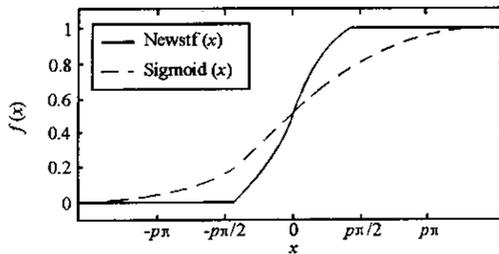


图 2 新的转移函数与 Sigmoid 函数的函数曲线

Fig.2 Function curves of new transfer function and Sigmoid function

4 实例应用及比较

心肌梗死是冠心病的严重临床症状,其定位识别诊断是医学数据自动处理中的一个典型的模式识别问题,本文采用新的模糊神经网络分类器进行实现。

为确立网络结构,首先要根据有关的医学知识总结出所有的诊断规则。例如,下面给出的就是在心肌梗死诊断中用到的而且是比较重要的两条诊断规则。

规则(1) if I 导联的 Q 波振幅 $> 0.2\text{mV}$ or $Q/R > 25\%$ or Q 波时限 $> 0.04\text{s}$
then I 导联的 Q 波异常

规则(2) if V1 出现异常 Q 波 and V2 出现异常 Q 波 and V3 出现异常 Q 波
then 前间壁心肌梗死

诊断中用到 11 个导联,根据需要在每个导联选取 Q 波的振幅、时限和 Q/R 比三个特征中的部分或全部作为输入特征参数。根据 Q 波的特点,将每个输入特征模糊化为“大”和“小”两个概念。隶属函数就选用本文所提出的新的 S 型转移函数,通过使用不同的权值和阈值来近似所需的上升或下降的形态。诊断结果要通过两级推理实现,一级推理中有 22 条规则,二级推理中有 7 条规则,所以推理层设为两层,节点个数分别为 22 和 7。心肌梗死有 6 类病态结果,再加上正常状态的结果,输出层节点数为 7。最后,我们得到的网络结构为 $30 - 60 - 22 - 7 - 7$ 。

实验运用了已确诊的 50 例心肌梗死的数据,每次随机选取 30 个病例组成一个样本组对神经网络进行训练,再用剩余的 20 例进行检验。与采用误差平方和准则及 Sigmoid 转移函数的网络进行比较,结果如表 1 所示。

表1 网络性能比较

Tab.1 Performance comparison of networks

网络类型	样本组 1		样本组 2		样本组 3	
	训练次数	正确识别率	训练次数	正确识别率	训练次数	正确识别率
传统模糊 BP 分类器	3918	58%	4217	57%	4236	61%
新的模糊神经分类器	879	80%	932	81%	892	87%

从表1的网络性能比较实验结果可以看出,无论是从网络训练的收敛速度看,还是从分类的正确率来看,基于交叉熵准则和新的转移函数的模糊神经网络分类器都优于传统的模糊BP神经网络分类器的性能。

5 结论

提出了基于交叉熵准则和新的S型转移函数的模糊神经网络分类器。基于交叉熵准则的BP算法与基于误差平方和准则的BP算法相比,不易产生振荡,收敛速度更快,不易陷入局部极小点,而且交叉熵函数也更适于作为网络分类器的目标函数。新的S型转移函数的优点在于可以取到0、1值,又具有比Sigmoid函数更陡峭的函数曲线,具有较快的网络学习速度。而且,网络的每一个节点、每一个连接都是根据诊断规则建立的,具有明显的物理意义,网络不再是个“黑箱”,可以对网络的输出做出合理的物理解释。最后,通过在心肌梗死定位诊断中的实际应用,验证了本文提出的基于交叉熵准则和新的S型转移函数的模糊神经网络分类器的有效性。

参考文献:

- [1] 阎平凡,张长水.神经网络与模拟进化计算[M].北京:清华大学出版社,2000.
- [2] 袁曾任.神经网络原理及其应用[M].北京:清华大学出版社,1999.
- [3] Sankar K P, Sushmita M. Multi-layer Perceptron, Fuzzy Sets and Classification[J]. IEEE Trans on Neural Networks, 1992, 3(5): 683 - 697.
- [4] 孙泽行,胡绍海,张思东,等. BP网络中的熵函数准则(英文)[J]. 北方交通大学学报, 1997, 21(5): 543 - 547.
- [5] 孙即祥.现代模式识别[M].长沙:国防科技大学出版社,2002.

