

一种多分辨率组合的模糊神经网络分类器

陈曦, 靳东明, 李志坚
(清华大学微电子学研究所, 北京 100084)

摘要: 提出一种多分辨率组合的分类器(MRCC)模型和相应的学习算法, 发展了 Simpson 的模糊最小最大神经网络(FMM)方法. 它克服了原始模型的几个缺点: 训练结果不依赖于训练样本出现的次序, 超盒扩张不受一个固定的最大尺度限制. 和原始模型的超盒相比较, 新模型的超盒中引入了一个参数表示超盒对于训练样本的分类正确率, 称为超盒置信度. 新的学习算法假设样本在一定尺度下均匀分布, 从而能够在线调整超盒置信度参数. 新的学习算法采用多分辨率组合的方法, 消除了原始算法中选择超盒最大尺度限制参数的困难. 实验表明, MRCC模型与原始FMM模型相比, 分类性能更好, 学习算法的自适应能力更强, 建立的模糊超盒数更少, 并行处理能力更强.

关键词: 模式分类; 模糊最小最大神经网络; 递归训练; 多分辨率组合; 局部均匀分布; 超盒置信度
中图分类号: TN18 **文献标识码:** A **文章编号:** 0372-2112 (2002) 06-0928-06

A New Kind of Fuzzy Neural Network Classifier

CHEN Xi, JIN Dong-ming, LI Zhi-jian

(Institute of Microelectronics Tsinghua University, Beijing 100084, China)

Abstract: A Multi-Resolution Combined Classifier (MRCC) model and its learning algorithm are proposed. The MRCC is a modification of Simpson's original Fuzzy Min-Max (FMM) Neural Network Classifier. It overcomes some undesired properties of the original model: specifically, training results do not depend heavily on pattern presentation order and hyper-box expansion is not limited by a fixed maximum size. Compared with the original model, a new parameter called the reliability of hyper-box is introduced in MRCC, which presents the correct classification rate of a single hyper-box. A locally uniform distribution hypothesis is also introduced, so that the reliability of hyper-box parameter can be adjusted on-line. In addition, the multi-resolution combination method alleviates the effort to select an optimal parameter for the maximum size of a hyper-box in the original algorithm. Experiments were made following some recent evaluation criteria known in literature, and show that compared with original model, the MRCC model has better classification performance, more adaptive learning ability and creates less hyper-boxes.

Key words: pattern classification; fuzzy min-max neural network; recursive training; multi-resolution combination; locally uniform distribution; hyper-box reliability

1 引言

分类器是模式识别, 信号检测, 控制系统和预测等许多工程问题的重要组成部分. 模糊集合理论和神经网络理论是处理模式分类问题常用的方法. 在模式分类问题中, 类别和模式特征的定义往往存在模糊性, 所以模糊集合理论适合处理分类问题. 模糊逻辑系统(FLS)用 IF-THEN 形式的模糊规则表示知识. 人们容易理解和调整系统的工作过程. 神经网络理论通常处理分类和统计回归问题. 神经网络由许多简单的处理单元相互连接构成. 通过适当的学习算法, 神经网络系统的参数能够根据输入输出数据建立和优化. 模糊逻辑和神经网络在处理模式分类问题时各有各的优点和缺点. 模糊逻辑系统根据模糊规则把模式空间划分成相互重叠的模糊区域, 不适合处理输入较多的问题. 另外, 如何获得适当的模糊规则建立系

统也是一个复杂的问题. 神经网络具有学习能力, 但是学习过程相对缓慢. 另外, 分析和调整通过学习过程建立的神经网络往往比较困难.

结合模糊逻辑和神经网络的优点, 使新的模型既具有神经网络的学习能力, 优化能力和连接结构, 也具有模糊系统 IF-THEN 形式的规则表示. 模糊神经网络是一种令人关注的模型. Simpson 提出的模糊最小最大神经网络(FMM)^[1]是一种采用超盒隶属度函数的模糊神经网络, 可以应用于模式分类, 函数逼近和聚类分析等问题. 最近得到了广泛的研究, 并且在字符识别和检测等领域得到应用. 但是, FMM模型和学习算法存在两个重要问题: 难于针对特定问题确定超盒尺度限制参数, 分类性能强烈依赖输入模式出现的次序. 因此, 本文发展了 FMM的思想, 提出一种多分辨率组合的模糊最小最大神经网络分类器模型(MRCC).

收稿日期: 2001-05-31; 修回日期: 2001-12-05

基金项目: 高等学校博士学科点专项科研基金(No. 98000338)

2 模糊逻辑系统的基本结构和模糊分类器

一个模糊逻辑系统如图 1,把非模糊的输入 x 映射到非模糊的输出 y .它由四个部分组成:模糊规则库,模糊化模块,推理模块和去模糊化模块.规则可以从专家知识获得,也可以从数值数据中提取.无论哪一种情况,规则都可以用如下 IF-THEN 的形式表示

$$R^l: \text{IF } x_1 \text{ is } F_1 \text{ and } x_2 \text{ is } F_2 \text{ and } \dots \text{ and } x_n \text{ is } F_n \text{ THEN } y \text{ is } G^l \quad (1)$$

其中 F_l 和 G^l 是模糊集合, $x = (x_1, x_2, \dots, x_n)$ 和 y 分别是输入和输出的语言变量(特征名称和类别名称), $l = 1, 2, \dots, L$.

模糊化模块把非模糊数值映射成模糊数.推理模块把模糊数映射成模糊数,它控制规则组合的方式.通常要求 FLS 的输出是非模糊数值,去模糊化模块把模糊数映射成非模糊数值.

在许多模糊系统和模糊神经网络中,输入和输出空间被划分成栅格形式.每一个栅格定义成一个模糊区域,栅格之间的重叠区域使隶属度的输出曲面平滑连续.虽然在学习过程中,隶属度函数的形状和位置可以被改变,但它们本质上还是栅格.栅格划分使得模糊系统的软硬件实现比较容易,但是随着输入输出变量的增加,划分的栅格数目呈组合增加,对于存储器和硬件的需求巨大,使系统无法实现.这种情况对学习过程的影响更大,因为精细的空间划分需要更多的训练样本,否则就无法充分学习.因此,复杂系统需要更加灵活的模式空间划分方法.FMM 采用模糊超盒划分模式空间,根据已知的输入输出数据建立划分,并且在线调整有关参数.超盒划分根据输入输出数据建立,和栅格划分相比更加灵活,减少了不必要的模式空间划分,降低了模糊逻辑系统的规模.

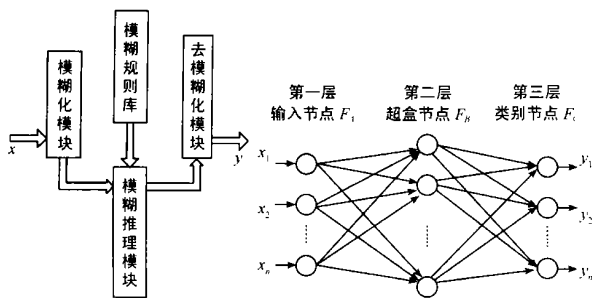


图 1 模糊逻辑系统

图 2 模糊最小最大神经网络结构

3 FMM 神经网络分类器

3.1 概述

模糊最小最大分类神经网络,采用超盒模糊集对模式空间进行划分.一个超盒表示 n 维模式空间中的一个区域,包含在超盒中的模式隶属于相同的类别.超盒由它在空间中的最小点和最大点表示,并且对应一个模糊隶属度函数.最小最大点和超盒隶属度函数描述了一个模糊集合.在分类问题中,表示相同类别的超盒模糊集合的并集构成表示表示一个类别的模糊集合.这些描述建立的模型具有神经网络的连接结构,如图 2,输入层 F_A 表示特征节点,隐含层 F_B 表示超盒节点,

即所建立的超盒模糊集合.输出层 F_C 表示类别节点.连接 F_A 和 F_B 层的是矢量 V 和 W ,包含超盒最小点和最大点的信息.连接 F_B 和 F_C 层的是二进制值,表示超盒的类别.这些连接通过离线学习算法建立.

FMM 的学习算法包括四个步骤:初始化,扩张,重叠检测,和收缩,对应每一个训练模式,重复执行后三个步骤,直到所有训练模式被网络学习一遍.

3.2 FMM 神经网络分类器的缺点和改进

FMM 模型和学习算法存在两个重要问题:难于针对问题确定适当的超盒最大尺度限制,分类性能强烈依赖输入模式出现的次序.超盒最大尺度限制的大小影响分析整个输入空间的分辨率,这对分类器性能的影响很大.对于这些问题,有各种改进的方法.S. Abe (1995)通过递归的划分超盒重叠区域,从数值数据中提取模糊规则.M. Meneganti (1998)提出一种新的学习算法,初始化超盒数等于待分类模式的类别数,然后通过分解和划分超盒的重叠区域等操作建立分类器模型.B. Gabrys (2000)^[2]在进行推广 FMM 模型研究的同时,提出逐渐减小 α 的取值,通过比较选择最佳值的方法.A. Rizzi (2000)提出采用超盒剪切的方法建立模糊最小最大网络模型.以上这些方法或者选择一个固定的值作为所建立超盒的最大尺度,或者找到超盒的重叠区域进行进一步的划分.但是,一个固定的超盒的最大尺度往往不能在模式空间的每一个部分都得到优化的结果,划分重叠区域的方法容易建立过多的超盒.我们在这些工作的基础上进行深入分析,提出适当的处理方法.

实际上,分类器训练得到的分类性能强烈依赖输入模式出现的次序这一点,部分是因为训练过程中超盒之间的竞争.相同类别超盒之间的竞争表现为只有对于输入模式隶属度值最大的超盒执行扩张包含输入样本;不同类别的超盒之间也存在竞争,表现为 FMM 模型规定这些超盒之间不能发生重叠,在每一步可能产生重叠的步骤结束以后,都要进行重叠检测,如果发现重叠,还需要通过超盒收缩消除重叠.这些都是因为设定每一个超盒内部区域中的模式隶属于这个超盒表示的类别的隶属度值都等于 1 造成的.如果两个表示不同类别的超盒相互重叠,那么重叠区域内部的模式隶属于两个类别的隶属度都等于 1,这样就无法对这个区域中的模式进行分类.这样为了避免冲突,输入的新模式就可能造成已存在超盒收缩的必要,而较早建立的超盒在训练过程中被要求收缩的概率较大.所以就造成了训练结果严重依赖模式输入的次序.我们在模型中引入超盒置信度的概念,允许表示不同类别的超盒相互重叠,从而消除表示不同类别的超盒在学习过程中的竞争.这样,就减小训练模式输入的次序对于分类器训练结果的影响,同时也简化了学习算法.

虽然在模型中引入超盒置信度的概念很大程度上简化了学习算法,但是确定每一个超盒的超盒置信度需要训练样本两次通过分类器,第二次数据通过仅仅执行对超盒性能的评价,这使算法的效率降低,尤其对于训练样本数较大的情况.为了处理这一个问题,在模型中引入了局部均匀分布假设,认为在每一个建立的超盒内部,各个类别的训练模式和全部训

训练模式都是均匀分布的. 引入局部均匀分布假设后, 新的模型和 FMM 的学习算法相似, 全部训练样本的学习只进行一次通过. 但是新模型的学习算法比 FMM 的学习算法简单, 当一个训练样本作用于这种模型的时候, 只需要选择一个超盒进行超盒扩张 (或者新建一个超盒), 并且调整所有已经建立的超盒的可信度. 所以, 新模型的学习算法具有良好的并行执行能力.

在此基础上, 本文进一步提出一种多分辨率超盒组合的模糊最小最大网络模型及其学习算法. 这种方法在每一次对全部训练样本的学习过程以后, 对所有存在的超盒的分类性能进行评价, 删除“不好”的超盒, 保留“良好”的超盒. 然后减小的取值, 再次进行训练, 直到分类器的性能满足要求, 或者达到的最小允许取值. 对于训练样本数量有限的问题, 采用全部训练样本循环训练.

4 MRCC 模型

4.1 基本定义

MRCC 模型处理输入模式的特征采用实数表示的分类问题, 训练样本集合是如下形式的输入输出有序对 (模式 - 类别对)

$$\{X_h, d_h\} \quad h=1, \dots, H \quad (2)$$

其中 $X_h = \{x_{h1}, x_{h2}, \dots, x_{hn}\}$ 是第 h 个输入模式, 不失一般性, 假设 X_h 是归一化的, 即 $X_h \in [0, 1]$, I_n 是单位长度. d_h ($1, 2, \dots, K$) 是第 h 个输入模式的类别标记, 用正整数表示.

4.2 模糊超盒隶属度函数

在模糊逻辑系统中, 经常采用梯形隶属度函数, 因为它能够简明直观的表达专家的知识. 在模糊最小最大分类器中, 模糊超盒隶属度函数扮演了重要的角色. 主要根据输入模式隶属于每一个超盒的程度, 它最终决定了一个出现的输入模式应该属于哪一个类别.

一个最小最大超盒由它在空间中的最小点和最大点描述, 依次表示为向量 V_j 和 W_j , 用隶属度函数 b_j 确定空间任意一点 X_h 隶属于这个超盒的隶属度值. 通过比较 X_h 隶属于不同类别超盒的隶属度值对 X_h 进行分类. 每一个超盒模糊集合 B_j 定义为一个有序集合

$$B_j = \{X_h, V_j, W_j, P_j, b_j(X_h, V_j, W_j, P_j)\} \quad (3)$$

对于全部 $h=1, 2, \dots, H$, 其中 X_h 是第 h 个输入模式, $V_j = (V_{j1}, V_{j2}, \dots, V_{jn})$ 是第 j 个超盒 B_j 的最小点, $W_j = (W_{j1}, W_{j2}, \dots, W_{jn})$ 是第 j 个超盒 B_j 的最大点, P_j 是超盒置信度参数, 表示一个模糊超盒分类性能的参数, 分类正确概率 p_j , 定义为超盒内部分类正确的训练样本数与超盒内部全部训练样本的总数的比值. 这个参数也可以理解为模糊超盒分类的置信度.

$$p_j = \frac{\text{超盒 } B_j \text{ 中分类正确的训练模式数 } (N^*)}{\text{超盒 } B_j \text{ 中全部训练模式的总数 } (N)} \quad (4)$$

第 j 个盒形 B_j 的隶属度函数是 $b_j(X_h, V_j, W_j, P_j)$, 满足 $0 \leq b_j \leq 1$.

$$b_j(X_h) = \min_{i=1 \dots n} (\min([1 - f(x_{hi} - w_{ji}, r_i)], [1 - f(v_{ji} - x_{hi}, r_i)])) \quad (5)$$

其中, $f(z, r)$ 是两个参数的斜坡阈值函数

$$f(z, r) = \begin{cases} p_j, & \text{if } zr > p_j \\ zr, & \text{if } 0 \leq zr \leq p_j \\ 0, & \text{if } zr < 0 \end{cases} \quad (6)$$

其中 x_{hi}, v_{ji}, w_{ji} 是矢量 X_h, V_j, W_j 的第 i 个单元, r 是灵敏度参数, 调整隶属度函数随 X_h 和 B_j 的距离增加而减小的速度. 简单的讲, 这个隶属度函数计算输入模式在各个维度上的隶属度值的最小值. 一维的隶属度函数如图 3 所示. 第 k 种模式类 C_k 的模糊集合定义为

$$C_k = \bigcap_j B_j \quad (7)$$

表示相同类别的超盒之间的重叠不会使分类过程出现问题. 对于表示不同类别的超盒的重叠部分, FMM 模型无法正确分类. MRCC 模型引入超盒可信度参数, 只要重叠的超盒的可信度参数值不同, 那么重叠部分的分类与可信度最高的超盒表示的类别相同. 所以只要超盒可信度值的差异足够大, 或者说, 每一个超盒中包含的训练样本数足够大, 就不需要消除表示不同类别的超盒之间的重叠.

4.3 局部均匀分布假设

在模型中引入超盒置信度的概念很大程度上简化了学习算法, 但是, 确定每一个超盒的超盒置信度需要训练样本两次通过分类器, 而第二次数据通过仅仅执行对超盒性能的评价, 这使得算法的效率降低, 尤其对于训练样本数很大的情况. 为了处理这个问题, 本文引入局部均匀分布假设.

局部均匀分布假设 (LUD) 是指在模式空间足够小的区域内, 各个类别的模式以及全部训练模式都是均匀分布的. 对于 FMM 模型而言, 假设这个足够小的区域就是由超盒最大尺度限制规定的多维空间区域. 简单的讲, 就是假定在每一个建立的超盒内部, 各个类别的训练模式和全部训练模式都是均匀分布的.

为了表示方便, 定义模式的空间密度为

$$= \frac{\text{模式子空间中的模式数 } (N)}{\text{模式子空间的体积 } (V)} \quad (8)$$

按照局部均匀分布假设, 在超盒 B_j 中模式的空间密度为一个常数, 于是, 超盒 B_j 的分类正确概率 p_j 可以改写为

$$p_j = \frac{N^*}{N} = \frac{N^*}{V} \quad (9)$$

在下面的讨论中可以看到, 经过这样的变换, 使得 p_j 可以在一个训练样本输入时进行在线调整.

4.4 MRCC 网络模型分类性能的评价

原始 FMM 的学习算法是一个全部训练样本只进行一次通过的学习算法, 当一个训练样本作用于 FMM 模型的时候, 模型对于这个样本的学习依次进行: 超盒扩张 (或者新建), 重叠检测, 超盒收缩 (如果需要) 等步骤. 全部训练模式依次对网络进行训练, 当每一个训练模式都使用过一次以后, 训练过程就结束了. 参数 r 控制了所建立超盒的最大尺度, 它在学习过程的开始就被设定, 并且在整个学习过程中保持不变. 但是, 至今并没有适当的算法确定 r 的优化值. 为了寻找这个参数的最佳值, 实际需要选择几个不同的 r 值对网络进行训练, 然后根据几次的分类错误率从中选择恰当的 r 值.

通过采用几组不同类型的数据对算法进行测试,我们发现网络训练过程中固定参数的值对性能和建立的超盒数有不好的影响。如果的值太大,会造成较大的错误分类率,尤其对于复杂、相互重叠的类别。另一方面,如果的值太小,会建立许多不必要的超盒,尤其对于集中、相互独立的模式通常属于同一个类别的情况。而且,建立过多的超盒使得分类器对于测试模式的分类能力降低。但是较小的取值有助于划分重叠的类别。

根据学习理论的基本原理,在对于训练集合的分类性能相同的条件下,结构复杂度最低的的网络的推广能力最好。根据这一原理,用于优化的目标函数包含两项,分别衡量网络分类的错误率和网络的复杂度,最优的网络结构能够最小化这样的目标函数:

$$F = (1 - \alpha) E + C \quad (10)$$

其中 E 是网络对于训练模式分类的错误率, C 是建立的超盒数与训练集中的模式总数之比,表征结构复杂度, $\alpha \in [0, 1]$ 是根据经验确定的权重。如果设置的值比较小,那么建立的网络结构就会比较复杂,对于训练模式的分类错误率也比较低,反之,如果设置的值比较大,那么建立的网络结构就会相对简单,对于训练模式的分类错误率也相对较高。在一般情况下,可以设置 $\alpha = 0.5$ 。

4.5 MRCC 的学习算法

前面提到,为了处理 FMM 模型中取值的任意性对网络性能的影响,MRCC 模型采用循环训练,多尺度超盒组合的方法,建立“训练-评价”结构的学习算法。具体方法是开始是用较大的值训练网络,在每一次训练结束以后,评价所建立的网络和每一个超盒的性能,其中,网络的性能由分类错误率 E 和结构复杂度 C 表示,超盒的性能由超盒可信度 p_j 表示。如果网络的性能已经满足要求,训练过程就结束了;如果网络的性能还没有达到要求,则删除可信度低于某一个阈值 p_{th} 的超盒。然后减小 α 的取值(如果需要),再次用全部训练模式重新训练网络,如此反复。在原始的算法中,训练过程在用全部训练模式对网络进行一次训练以后就结束了。采用多尺度组合的方法则需要定义训练的停止条件,满足这个条件训练过程就认为结束了。除了前面提到的网络性能满足要求,还需要人为规定一些对于模型的限制,为了防止一个超盒只包含一个模式的极端情况,规定超盒的最小允许尺度为 \min ; 为了防止循环训练的次数过多,导致网络学习的速度太慢,通过设置最大循环训练次数或者选择适当的减小步长,控制训练过程的结束。

4.5.1 初始化样本检测 检测输入的模式满足以下三种情况的哪一种:(1)输入模式包含在表示相同类别的超盒中。(2)存在表示相同类别的超盒,在满足超盒尺度限制条件下,能够经过扩张包含这个输入模式。(3)以上两种情况都不满足,需要根据这个输入模式建立一个新的超盒。

针对以上三种情况,相应进行以下操作:(a)不需要进行后面的学习步骤。(b)进行如下叙述的超盒扩张。(c)根据这个输入模式建立一个新的超盒 B_j ,实际上新的超盒是模式空间的一个点

$$V_j = X_h, W_j = X_h \quad (11)$$

4.5.2 超盒扩张 如果第 h 个输入模式 X_h 不包含在已经建立的同类别的超盒中,那么需要找到对应它的隶属度值最大的超盒 B_j ,并且超盒 B_j 满足超盒扩张准则。为了约束建立的超盒的最大尺度,规定超盒尺度满足如下限制:

$$\forall_{i=1, \dots, n} 0 \leq W_i - V_i \leq \quad (12)$$

其中, $\alpha \in [0, 1]$ 。这样分别限制了超盒每一个维度的尺度都小于 α 。根据超盒尺度限制准则,可以定义超盒扩张准则:

$$\forall_{i=1, \dots, n} (\max(W_{ji}, X_{hi}) - \min(V_{ji}, X_{hi})) \leq \quad (13)$$

对于输入模式 $\{X_h, D_h\}$, 如果一个超盒 B_j 满足上述扩张准则,那么

$$\text{if } \text{class}(B_j) = \begin{cases} D_h \Rightarrow \text{adjust}(B_j) \\ \text{else} \Rightarrow \text{select another } B_j \end{cases} \quad (14)$$

其中, $\text{adjust}(B_j)$ 是指如下操作

$$V_{ji}^{\text{new}} = \min(V_{ji}^{\text{old}}, X_{hi}) \quad \text{for } \forall i = 1, \dots, n \quad (15)$$

$$W_{ji}^{\text{new}} = \max(W_{ji}^{\text{old}}, X_{hi}) \quad \text{for } \forall i = 1, \dots, n$$

按照最小调整原则,只有一个超盒进行调整,但是进行调整的维度可能不止一个。

4.5.3 超盒可信度调整 对应输入的新训练样本,需要调整已经建立的各个超盒的可信度。根据训练样本和超盒的关系,存在以下几种情况:模式 X_h 不包含在超盒 B_j 内部,模式 X_h 包含在超盒 B_j 内部,以及超盒 B_j 经过超盒扩张包含模式 X_h 。另外,考虑模式 X_h 和超盒 B_j 的类别,下面具体讨论各种可能的情况和处理方法。(1)模式 X_h 不包含在超盒 B_j 内部:如果模式 X_h 不包含在超盒 B_j 内部,那么不需要调整超盒 B_j 的置信度参数 p_j 。(2)模式 X_h 包含在超盒 B_j 内部:如果模式 X_h 包含在超盒 B_j 内部,那么超盒 B_j 的可信度参数 p_j 需要根据这个输入模式进行调整。如果模式 X_h 的类别与超盒 B_j 表示的类别相同,那么超盒 B_j 的可信度参数 p_j 应该相应的有所提高

$$p_j^{\text{new}} = p_j^{\text{old}} \cdot \frac{N^* + 1}{N^* + p_j^{\text{old}}} \quad (16)$$

如果模式 X_h 的类别与超盒 B_j 表示的类别不相同,那么超盒 B_j 的可信度参数 p_j 应该相应的有所降低

$$p_j^{\text{new}} = p_j^{\text{old}} \cdot \frac{N^*}{N^* + p_j^{\text{old}}} \quad (17)$$

(3)超盒 B_j 经过超盒扩张包含模式 X_h :如果超盒 B_j 经过超盒扩张包含模式 X_h ,那么超盒 B_j 的可信度参数 p_j 需要根据这个输入模式进行调整。因为超盒扩张是学习算法中唯一改变超盒形状的操作,所以在理想状态下,超盒扩张所增大的区域中除了这个输入的模式以外,不包含其它相同类别的模式。另一方面,对于与超盒表示的类别不同的模式,按照局部均匀分布假设,这些模式的空间密度在扩张前后的超盒中保持不变,如图 4 所示。

因此,可以得到扩张以后超盒 B_j 的可信度参数

$$p_j^{\text{new}} = A_p \cdot p_j^{\text{old}} \quad (18)$$

其中

$$A_p = \frac{N^* + 1}{(N^* + 1) \cdot p_j^{\text{old}} + A \cdot N^* \cdot (1 - p_j^{\text{old}})} \quad (19)$$

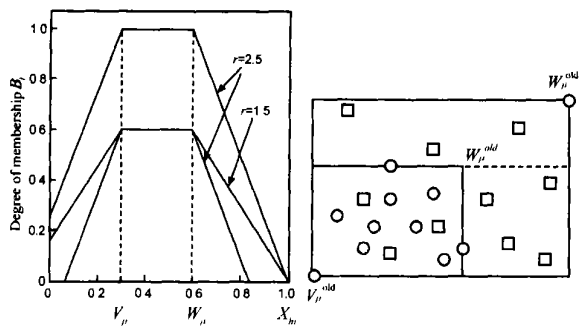


图 3 超盒 B_j 的一维(第 i 维)隶属度函数

图 4 各类样本在超盒 B_j 扩张过程中的局部均匀分布假设。

· 输入模式 X_n , 类别 1;
· 输入模式 X_n , 类别 2

因为在实际训练过程中, 当一个超盒内部的训练样本总数较少时, 用这些样本计算样本的空间密度和实际情况相差较大, 为了防止这种情况对训练过程的影响, 规定超盒可信度改变因子 A_p 应该接近于 1.

$$A_p = \begin{cases} 0.5, & \text{if } A_p < 0.5 \\ A_p, & \text{if } 0.5 \leq A_p \leq 2 \\ 2, & \text{if } 2 < A_p \end{cases} \quad (20)$$

其中, 主要参数是 A_p 的最小值设为 0.5, 它的含义是当样本太少而不足以估计总体分布的情况下, 认为所有样本是均匀分布的. 对每一个输入的训练模式, 依次执行以上三步操作. 当全部训练模式被用来训练分类器一遍以后, 训练过程就结束了.

4.5.4 评价网络和模糊超盒的性能, 检测训练停止条件 检测两个训练停止条件: (1) 在全部训练模式进行训练以后, 目标函数 F 的值达到预先定义的要求 F_{\min} . (2) ϕ 值的减小达到人为定义的最小值 ϕ_{\min} , 这样可以防止一个超盒只记录一个模式的极端情况. 如果以上两个训练停止条件之一满足, 那么训练过程就结束了. 否则, 按照如下方法减小 ϕ 的取值

$$\phi_{\text{new}} = \phi_{\text{old}} \cdot \phi \quad \text{或} \quad \phi_{\text{new}} = \phi_{\text{old}} - \phi \quad (21)$$

其中 ϕ 表示 ϕ 减小的速度 ($0 < \phi < 1$), ϕ_{old} 是 ϕ 减小的尺度量 ($0 < \phi_{\text{old}} < 1$). 并且删除可信度低于阈值的超盒, 保留其它已经建立的超盒. 然后转向 4.5.1 重新进行新的一遍对于全部训练模式的学习. 和第一遍训练不同的是, 超盒最大尺度限制参数减小了, 而且已经存在一些分类性能达到阈值的超盒.

5 实验结果

为了检验新模型的性能, 采用合成数据集和标准数据集对 MRCC 的性能进行检验, 包括两类二维高斯分布, IRIS, wine, ionosphere, sonar, PIMA 和 BUPA^[4] 数据集. 我们的试验局限于比较原始 FMM 和 MRCC 模型, 而与其它统计方法, 模糊技术和神经网络模型比较, 可以参考其它文献. 为了得到的比较结果有意义, 比较算法时采用相同的实验环境, 包括对训练模式集合和测试模式集合的划分相同, 并且输入模式的次序也相同.

5.1 合成数据集

本实验使用的训练样本集合产生于两类二维正态分布的数据模型, S. Haykin (1999)^[3] 使用这个问题研究 MLP 网络和 RBF 网络用作的分类器的性能. 本实验的目的是区分两类二维正态分布的模式, 分别记为 A 和 B , 它们的类条件概率密度函数 ($p. d. f.$) 如下

$$f(x|w_A) = \frac{1}{2 \cdot \dots \cdot 2 \cdot A} \exp \left\{ - \frac{x - \mu_A}{2 \cdot A} \right\}$$

$$\mu_A = [0 \quad 0]^T, \quad A = 1$$

$$f(x|w_B) = \frac{1}{2 \cdot \dots \cdot 2 \cdot B} \exp \left\{ - \frac{x - \mu_B}{2 \cdot B} \right\}$$

$$\mu_B = [2 \quad 0]^T, \quad B = 2 \quad (22)$$

其中 w_{\otimes} 表示 x 属于类别 \otimes , \cdot 是 R^2 空间中的标准欧式距离, 两类数据出现的概率相同, $P(w_A) = P(w_B) = 0.5$. $f(x|w_A)$ 和 $f(x|w_B)$ 分别是两个独立的一维随机变量 ($r. v.$.) 的联合概率密度.

根据贝叶斯理论, 一个两类分类问题最佳决策边界可以通过最大化分类器对于一个的输入模式 $x = [x_1, x_2]^T$ 的决策的后验概率得到. 通过计算可以得到贝叶斯分类器应用于这个分类问题的平均错误概率, 即错分率 $P_e = 0.185$, 正确分类的概率是 $P_c = 1 - P_e = 0.815$. 对于这个模式分类问题, S. Haykin (1999)^[3] 仔细研究了 BP 算法学习的 MLP 网络和规则化的 RBF 网络的分类能力. 我们采用这个问题检验原始的 FMM 模型和 MRCC 模型, 每一类数据随机产生 500 个模式, 两类数据一共 1000 个模式. 从模式集合中随机选取 50% 的模式组成训练模式集合, 剩下 50% 的模式组成测试模式集合. 这种划分方法能够比较客观的体现模型对没有出现在训练过程中的模式进行分类的能力, 即模型的推广能力. 实验结果 MRCC 模型对测试样本进行分类的正确率是 78.6%, 优于原始 FMM 模型得到的最好结果 76.4%. 并且, MRCC 模型在训练过程中建立的超盒数为 28 个, 小于获得最佳结果时原始 FMM 模型所建立的超盒数 38 个.

MRCC 模型在训练过程中假设在超盒尺度最大限制的局部区域内, 各类模式都是均匀分布的. 这个假设使得算法能够在线调整超盒置信度, 从而简化训练算法. 我们用上述 1000 个训练模式训练 MRCC 模型, 得到 11 个超盒, 我们比较通过训练得到的超盒置信度和根据定义得到的超盒置信度.

表 1 超盒置信度检验局部均匀分布假设

| 超盒编号 | 假设 | 定义 | 超盒编号 | 假设 | 定义 |
|------|--------|--------|------|--------|--------|
| 1 | 0.5070 | 0.5057 | 7 | 0.9874 | 0.9910 |
| 2 | 0.7278 | 0.7300 | 8 | 0.6640 | 0.6418 |
| 3 | 0.2331 | 0.2337 | 9 | 1.0000 | 1.0000 |
| 4 | 0.9280 | 0.9337 | 10 | 0.4792 | 0.5977 |
| 5 | 0.3476 | 0.3985 | 11 | 1.0000 | 0.6667 |
| 6 | 0.6326 | 0.6503 | - | - | - |

注: 假设 —— 根据 LUD 假设通过训练直接得到的超盒置信度
定义 —— 训练完成以后根据定义计算得到的超盒置信度参数
从上述结果表明根据局部均匀分布假设在线推算出的超盒置信度和实际情况吻合较好, 其中个别数值差别较大是因

为该超盒中得样本数较少所致. 考虑到超盒置信度之间的相对大小决定分类结果, 所以有理由在训练算法中采用局部均匀分布假设.

5.2 标准数据集超盒

本实验采用若干常用的真实数据集检验和比较 MRCC 模型和原始 FMM 模型的性能. 下面列出的有关原始 FMM 模型的结果, 是在不同训练参数条件下, 经过测试得到的最佳结果. 这里采用的实验方法如下: 前三个数据集划分成训练集合占全部的 50%, 测试集合是剩下的 50%; 后三个数据集划分成训练集合占全部的 75%, 测试集合是剩下的 25%. 下面给出的原始 FMM 模型的结果是选择不同训练参数的值训练得到的模型中, 在测试集合上表现分类正确率最高的一个. 而 MRCC 模型的结果是一次实验得到的结果.

表 2 MRCC 模型和原始 FMM 模型比较

| 数据集 | 原始 FMM 模型 | | | MRCC 模型 | | | |
|-------|-----------|---------|---------|---------|---------|---------|----|
| | NL/L | NT/T | NB | NL/L | NT/T | NB | |
| IRIS | 0.14 | 75/75 | 72/75 | 29 | 75/75 | 73/75 | 3 |
| WINE | 0.27 | 88/90 | 86/88 | 51 | 90/90 | 88/88 | 8 |
| IONO | 0.20 | 156/176 | 152/175 | 120 | 175/176 | 165/175 | 38 |
| SONAR | 0.24 | 133/157 | 36/51 | 136 | 157/157 | 36/51 | 17 |
| PIMA | 0.44 | 385/576 | 130/192 | 44 | 475/576 | 136/192 | 10 |
| BUPA | 0.12 | 231/259 | 54/86 | 138 | 254/259 | 57/86 | 99 |

注: NL/L —— 分类正确的训练模式/ 训练模式

NT/T —— 分类正确的测试模式/ 测试模式

NB —— 建立的超盒单元数

从上面的结果看到, 在相同条件下 MRCC 模型对于测试模式分类的正确率高于原始 FMM 模型获得的最佳结果, 并且 MRCC 模型建立的超盒数明显少于原始 FMM 模型建立的超盒数.

6 结论

学习分类器最重要的性能是它的推广能力, 对于 FMM 分类器模型而言, 推广能力主要决定于训练过程中建立的超盒的大小和位置, 对于 MRCC 模型还包括超盒置信度参数. MRCC 模型通过引入超盒置信度参数, 采用多分辨率组合的方法, 不仅简化了训练算法, 而且降低了所建立模型的复杂度, 提高了分类器的推广能力.

我们还着重考虑了训练过程的另一个重要特性, 即训练的自动化程度. MRCC 模型假设模式局部均匀分布, 使之保持了原始 FMM 模型具有一次通过的训练能力. 在训练开始之前, 用户只需要给出和期望的分类器性能相关的参数, 而无需给出其它对训练结果影响较大的关键参数. 因此, MRCC 模型的训练过程具有较高的自动化程度.

实验表明, MRCC 模型和原始 FMM 模型相比, 训练过程模型建立的超盒数较少, 分类推广能力更强, 并且训练的自动化程度更高.

参考文献:

- [1] P K Simpson. Fuzzy min-max neural networks-Part 1: classification [J]. IEEE Trans. on Neural Networks, 1992, 3(5): 776 - 786.
- [2] B Gabrys, A Bargiela. General fuzzy min-max neural network for clustering and classification [J]. IEEE Trans. Neural Networks, 2000, 11(3): 769 - 783.
- [3] S Haykin. Neural Networks: A Comprehensive Foundation, 2nd Edition [M]. Englewood Cliffs, NJ: Prentice-Hall, 1999.
- [4] C Blake, E Keogh, C J Merz. UCI repository of machine learning databases [EB/OL]. <http://www.ics.uci.edu/~mllearn/MLRepository.html>. 1998.

作者简介:



陈 曦 男, 1975 年 7 月出生于上海市, 清华大学微电子所博士研究生, 1998 年毕业于清华大学电子工程系, 直读博士学位, 主要研究方向是模糊逻辑、神经网络、学习算法和硬件实现.



靳东明 男, 1942 年 2 月出生于天津市, 清华大学微电子所教授、博导, 1965 年毕业于清华大学无线电系, 并留校任教, 目前主要在多值逻辑、模糊控制和神经网络的硬件实现方面开展研究工作.