

基于 MATLAB 的 BP-LVQ 神经网络组合分类模型

钟淑瑛, 李陶深

(广西大学 计算机与电子信息学院, 广西南宁 530004)

摘 要:综合考虑神经网络分类误差率以及训练速率,文中从组合分类器结构出发,提出一种树形多层的 BP-LVQ 神经网络组合分类器模型。该组合分类器利用 BP 神经网络独立性以及自适应性解决了一般分类器难以不断学习和适应新攻击的问题,利用 LVQ 神经网络的竞争性将客观分类信息转变成使用者所定义类别。利用 MATLAB 神经网络工具箱对该 BP-LVQ 神经网络组合分类模型进行仿真实验,实验结果表明,该组合分类器的分类效率明显高于单一分类器。

关键词:组合分类器;BP 神经网络;LVQ 神经网络;MATLAB

中图分类号:TP183

文献标识码:A

文章编号:1005-3751(2006)02-0114-03

BP-LVQ Neural Networks Classifiers Combination
Model Based on MATLAB

ZHONG Shu-ying, LI Tao-shen

(College of Computer, Electronics and Information, Guangxi University, Nanning 530004, China)

Abstract: By considering the error rates and the training speed of neural networks, a hierarchical classifiers which is called as BP-LVQ neural network combination model is proposed in this paper. Combining with the capability of independency and self-adaptation of BP neural network, this combined classifier can solve the problem that normal classifiers can't keep learning and adapting new attack, and the problem that how to change the impersonal class into the category user defined by competitive position of LVQ neural network. The experimental results show that the tree-schema model can keep better performance than single classifier.

Key words: combined classifier; BP neural network; LVQ neural network; MATLAB

0 引言

传统的模式识别系统采用一个简单特征表达和特定的分类器算法来决定一个给定模式的类别问题,当类别数较多或是输入模式噪声干扰比较厉害时,这种识别系统并不能够具备很好的性能。将多个分类器组合起来对样本信息进行识别是提高分类效率的主要手段之一,目前主要组合方式有样本集划分法、属性选取法、类别编码法和组合模型法^[1]。一般的组合模型多为两层结构:第一层为多个不同的学习算法独立地对训练样本进行训练;第二层为一个分类器,对第一层中各分类器的输出进行某种组合,且多为线性组合。

文中提出一种树形多层组合模型:最下层叶子层包含若干个子分类器,其唯一输出为该子分类器针对各类别赋予样本信息的权值,将作为向上一层子分类器的输入。以此类推,第 N 层的输出作为第 $N-1$ 层的输入,最后一层

树根层输出表征某样本信息属于各个类别的可信度。在此基础上,构建了一种树形二层组合模型,由若干个 BP 神经网络(Back Propagation Neural Networks)作为最下层子分类器,由 LVQ 神经网络作为最上层分类组合器^[2~4],利用 MATLAB 的 NNTOL 神经网络工具箱进行实验,结果表明,树形多层组合模型分类效率明显高于单一分类器。

1 树形多分类器组合模型

假设样本共有 N 类,其类别由 $C_j (j = 1, 2, \dots, N)$ 表示;采用 S 层树形多分类器组合模型进行分类,分类器记为 $\text{Classifier}_{i,j}^s (i = 1, 2, \dots, I, s = 1, 2, \dots, S)$,当 $j = 1, s = S$ 时, $\text{Classifier}_{i,1}^S$ 为最下层叶子层中针对类别 C_1 的第 i 个

分类器。对于某个测试样本 $x = \begin{bmatrix} x_{11} \cdots x_{1j} \\ \vdots \\ x_{m1} \cdots x_{mj} \end{bmatrix}$, 其中 m

$= 1, 2, \dots, M$ 。每层中得到输出向量: $y^s = (\overline{y_1^s}; \overline{y_2^s}; \cdots; \overline{y_I^s})$, 其中 $\overline{y_1^s} = (y_{11}^s, y_{12}^s, \dots, y_{1j}^s), \dots, \overline{y_I^s} = (y_{I1}^s, y_{I2}^s, \dots, y_{Ij}^s)$ 。

定义 1: 对于矩阵 $y = (\overline{y_1}; \overline{y_2}; \cdots; \overline{y_n}) =$

收稿日期:2005-05-19

基金项目:广西“新世纪十百千人才工程”专项基金(桂人字 2001213 号);广西自然科学基金项目(桂科自 0229008)

作者简介:钟淑瑛(1979—),女,广西陆川人,硕士研究生,研究方向为数据挖掘、知识发现;李陶深,教授,研究方向为分布式数据库、人工智能、信息与网络安全。

$$\begin{pmatrix} y_{11}y_{12}\cdots y_{1j} \\ y_{21}y_{22}\cdots y_{2j} \\ \cdots \\ y_{i1}y_{i2}\cdots y_{ij} \end{pmatrix}, \text{ 存在转置函数 } T, T(y) = \begin{pmatrix} y_{11}y_{12}\cdots y_{1j} \\ y_{21}y_{22}\cdots y_{2j} \\ \cdots \\ y_{i1}y_{i2}\cdots y_{ij} \end{pmatrix}^T = \begin{pmatrix} y_{11}y_{21}\cdots y_{i1} \\ y_{12}y_{22}\cdots y_{i2} \\ \cdots \\ y_{1j}y_{2j}\cdots y_{ij} \end{pmatrix}$$

1.1 树形多分类器组合模型描述

树形多分类器组合模型的构造思想是:每一层都包含多个子分类器,该层的输出作为上一层的输入;以此类推,最后一层(即树根层)的输出为判断结果。图1给出了一个S层的树形多分类器组合模型,其中假设叶子层的输入样本相同,其输入层神经元数目为样本矩阵的行数 m ,输出层神经元数均为1;介于叶子层和树根层之间的层称为中间层,中间层分类器结构是相同的,其输入的神经元数目均为样本类别数 N ,输出的层神经元数均为1;树根层分类器的输入层神经元数为样本类别数 N ,输出层神经元数为 N 。这些层的输入和输出说明如下:

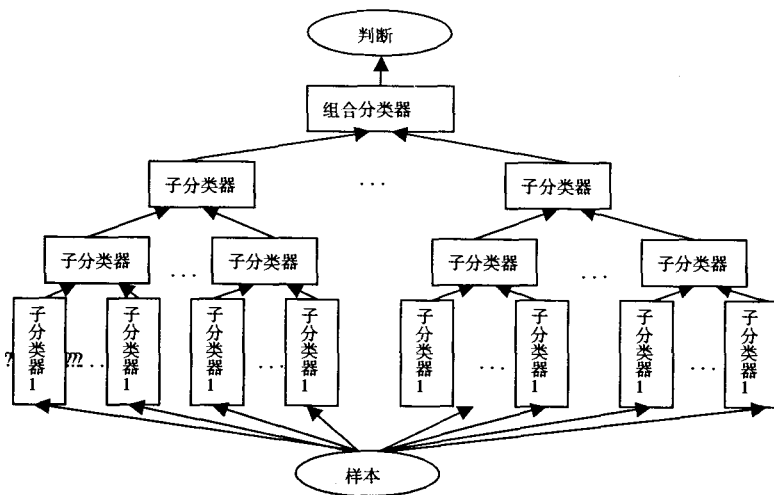


图1 树形多层分类器组合模型

(1) 叶子层分类器 $\text{Classifier}_{i,j}$ 的输出 $\bar{y}_{ij} = (y_{i1}, y_{i2}, \dots, y_{ij})$, 表征针对类别 C_j 所赋予样本 X 的权值, 该权值表示样本 X 属于类别 C_j 的概率, 且满足: $0 \leq y_{ij} \leq 1$ 。

(2) 中间层分类器 $\text{Classifier}_{i,j}$ 的输入为 $T(\bar{y}_1, \bar{y}_2, \dots, \bar{y}_i)$, $s \in [2, S-1]$; 输出为 $\bar{y}_s = (y_{s1}, y_{s2}, \dots, y_{sj})$ 表征针对类别 C_j 所赋予样本 X 的权值, 该权值表示样本 X 属于类别 C_j 的概率, 且满足: $0 \leq y_{sj} \leq 1$ 。

(3) 树根层分类器 Classifier^1 的输入为 $T(\bar{y}_1, \bar{y}_2, \dots, \bar{y}_i)$, 输出记为 $R = (\bar{R}_1, \bar{R}_2, \dots, \bar{R}_N)$, 其中, \bar{R}_N 表示样本 X 属于类别 N 的可信度。

1.2 树形多分类器组合模型特性

理想的分类器组合情况是各子分类器的训练结果相当准确且彼此间互不干扰, 有一定的独立性。这样的组合

应具备3个条件, 即不同的训练数据集, 不同的特征向量, 不同的训练参数或是不同的分类器结构^[5]。

文中提出的组合分类器属于基于类别和参数的混合型分类器, 即将叶子层子分类器按照测试样本的类别数目分成 N 个部分, 分别对 N 个训练数据集进行测试; 各类别样本的特征向量存在一定差异; 叶子层与非叶子层分类器的训练参数和结构均不相同。

图1所示的树形多层分类器组合模型兼顾了分类器的分类性能, 因此它也具有分类精度特性。

文中提出的树形多层分类器组合模型的组合分类器均为神经网络分类器, 因而具有神经网络分类器的特性: MSE (Mean Square Error)^[6]。MSE使用神经网络的输出与输入的均方误差来衡量网络性能, 即:

$$\text{MSE} = E(e^T \times e) = E((R - X)^T \times (R - X))$$

如果引入变量 k , k 为方差期望值被所得方差反复替换的次数, 即训练迭代次数。则MSE为:

$$\begin{aligned} \text{MSE} &= (R(k) - X(k))^T \times (R(k) - X(k)) \\ &= e^T(k) \times e(k) \end{aligned}$$

在MATLAB中, MSE是以下面的形式出现的:

$$\text{MSE} = \frac{\text{Performance}}{\text{Goal}}$$

2 BP-LVQ NN 组合模型

BP神经网络是一种单向传播的多层前向网络, 它除了有一个输入层和一个输出层之外, 还有一层或多层的隐层, 同层节点中没有任何耦合。输入信号从输入层节点一次传过各隐层节点, 然后传到输出节点, 每一层节点的输出只影响下一层节点的输出。BP神经网络具有的非线性映射能力保证其能够成功实现各种简单或复杂分类, 其并行结构加速了运算速度。另外, BP神经网络将信息分布式存储于连结权系数中, 使网络具有较高的容错性和鲁棒性。因此在目前的人工神经网络的实际应用中, 绝大部分的神经网络模型是采用BP神经网络及它的变化形式。然而, 由于BP神经网络输出层的结点数只能限制在一个较小的取值范围, 因而基于BP神经网络的分类器所不得不面临的一个问题就是如何才能做到分类的细致和多样性。

LVQ神经网络能够通过竞争性的隐层传递函数, 根据目标的类别将输入向量进行组合分类。

该网络由两层组成, 第一层为竞争层, 第二层为线性层。竞争层能够学习对输入向量的分类; 线性层将竞争层传来的分类信息转变成使用者所定义类别。LVQ神经网络的优势在于: 它不仅能够对线性输入数据进行分类, 而且还能处理多维的、甚至是含噪含干扰的数据量。实验已经证明, 只要在各个隐层存在足够的神经元, 那么目标输出的分类量将会得到相应的增加。

文中综合考虑了 BP 神经网络和 LVQ 神经网络的优缺点,利用 BP 神经网络作为第一层分类器, LVQ 神经网络作为第二层分类器,形成树形两层组合模型,结构如图 2 所示。

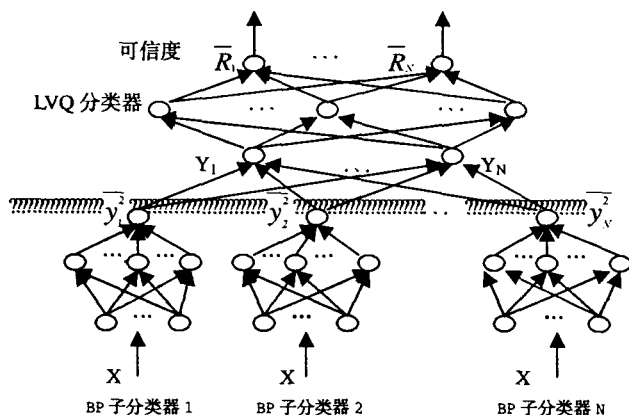


图 2 BP-LVQ NN 组合分类器模型

在该模型中, BP 神经网络分类器的输入样本为原始样本 X , 根据定义 1, LVQ 神经网络分类器的输入样本则为 $T(\bar{y}_1^2, \bar{y}_2^2, \dots, \bar{y}_N^2)$, 输出 $R = (\bar{R}_1, \bar{R}_2, \dots, \bar{R}_N)$ 。采用选择策略来挑出可信度估计值最大的一类作为对样本的判断^[7,8], 往往因为实际情况的需要不同而有失偏颇, 文中将样本的属于各个类别的可信度以向量的形式输出, 给予决策者更多的考虑空间, 从而根据实际情况做出最终判决。

3 实验结果分析

文中利用 UCI 机器学习库中 4 个不同领域的数据集^[9], 对树形 BP-LVQ 组合分类器进行实验, 并根据其组合特性进行分析。分别从 UCI 机器学习库中每个数据集抽取 2/3 的样本组成训练集对神经网络分类器进行训练, 其余 1/3 作为测试集测试分类性能。本实验中, 设定所需辨别的类别数目为 5 类, Goal 的取值为 $1e-006$, 当达到 Goal 时便停止训练神经网络; 由 MSE 值计算公式 $MSE = \frac{\text{Performance}}{\text{Goal}}$, 得出将组合分类器与简单 BP 神经网络分类器在各数据集下的 MSE 比较结果(如图 3 所示)。

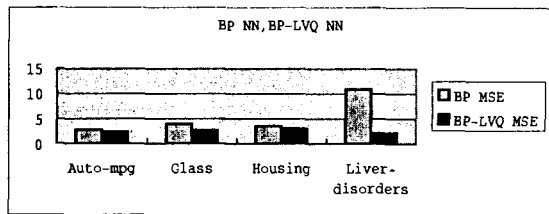


图 3 两种分类器的 MSE 值

笔者还对组合分类器与简单 BP 神经网络分类器作了训练集分类精度和测试集分类精度的实验比较, 比较的结果如表 1 所示。可以看出, BP-LVQ 树形组合分类器的 MSE 值偏小, 且训练集、测试集的分类精度略高于简单 BP NN 分类器。以上两点说明, 组合分类器具有良好的

组合特性和分类效率。

表 1 训练集分类精度与测试集分类精度

UCI 数据集	训练集分类精度 (%)		测试集分类精度 (%)	
	BP NN	组合分类器	BP NN	组合分类器
Auto-mpg	80.05	86.96	71.29	85.68
Glass	73.36	75.51	70.19	74.37
Housing	88.17	90.26	74.67	90.04
Liver-disorders	67.08	80.05	58.67	81.32

4 结束语

近年来, BP 神经网络研究领域的一个热点是将 BP 神经网络的自学习能力应用到分类系统之中, 对信息进行识别、分类。目前所实现的 BP 神经网络分类器普遍存在自训练时间过长、识别分类效率不高的问题。文中基于 MATLAB 神经网络工具箱, 提出一种树形多层组合分类器模型。该模型综合考虑了 BP 神经网络的独立性以及各种识别分类系统的样本信息的数据类型等因素, 利用 RP 训练算法取代传统的 LM 训练算法, 在不牺牲 BP 神经网络分类性能的基础上有效地缩短了训练时间; 结合 LVQ 神经网络分类组合器, 给出各子分类器的可信度, 并根据实际输出与期望输出的偏差, 及时调整对应的 BP 子分类器。经过对 BP-LVQ NN 两层组合分类器进行仿真实验, 结果表明, 该分类器在整个样本训练过程中表现了快速、自适应和易调整等特点, 能有效地提高分类效率。

参考文献:

- [1] Wolpert D. Stached Generalization[J]. Neural Network, 1992, 5(2): 241-260.
- [2] Hagan M T, Demuth H B, Beale M H. Neural Network Design [M]. Beijing: China Machine Press, 2002.
- [3] KANG HEE-JOONG. A Framework for Probabilistic Combination of Multiple Classifiers at an Abstract Level, Engng application[J]. Artificial Intellience, 1997, 10(4): 379-385.
- [4] Bilmes J A, Kirchhoff K. Generalized rules for combination and joint training of classifiers[J]. Pattern Analysis & Applications, 2003(6): 201-211.
- [5] Tolba A S, Abu-Rezq A N. Combined Classifiers for Invariant Face Recognition[J]. Pattern Analysis & Applications, 2000 (3): 289-302.
- [6] Kropas-Hughes C V, Rogers S K, Oxley M E. et al. Back-propagation of an Image Similarity Metric for Autoassociative Neural Networks[J]. Pattern Analysis & Applications, 2000 (3): 31-38.
- [7] Liu ChengLin. Classifier combination based on confidence transformation[J]. Pattern Recognition, 2005(38): 11-28.
- [8] Liu ChengLin, Hao Hongwei, Sako H. Confidence Transformation for Combining Classifiers[J]. Pattern Analysis & Applications, 2004(7): 2-17.
- [9] Blake C L, Merz C J. UCI Repository of machine learning databases[EB/OL]. <http://www.ics.uci.edu/mllearn/ML-Repository.html>, 1998.