

文章编号:1674-2869(2008)04-0114-04

一种面向特征选择的分类神经网络

杨海燕^{1,2}, 文一凭³

(1. 福建工程学院计算机与信息科学系, 福建 福州 350014; 2. 中南大学信息学院, 湖南 长沙 410075;
3. 湖南科技大学数学与计算科学学院, 湖南 湘潭 411201)

摘要:提出了一种面向特征选择的分类神经网络. 该网络对中心矢量和连接权同时学习, 其中心矢量是分类的中心, 而基于模糊隶属度的权表示特征的重要性, 根据权值大小进行特征取舍, 这样同时解决了模式识别中的模式分类和特征选择问题. 在 IRIS 数据集上的测试实验表明, 该网络能从原始特征中选择重要特征, 同时保持最大识别率.

关键词:特征选择; 分类; 中心矢量; 连接权

中图分类号:TP 391.41 **文献标识码:**A

0 引言

特征选择与分类是模式识别中两大问题. 在多数模式识别问题中, 先进行特征选择, 去掉对分类不相关的特征以及冗余特征, 从而减少数据处理时间以及数据存储空间, 然后再根据约简的特征进行分类.

近年来, 模糊逻辑与神经网络的融合方法研究得到了学术界的极大关注, 这是由于两者间具有互补性. 文献[1]使用模糊集来表示神经网络分类问题, 文献[2]引入均衡学习矢量量化(PLVQ)来解决传统学习矢量量化(LVQ)[3]中的初始化问题和规范化问题, 即有些权值因为输入矢量的分布, 可能无法在竞争中获胜, 因此就不能得到学习, 从而使神经网络单元没有被利用的问题.

下面对 PLVQ 及其中的 FLL 进行简介, 对提出的分类神经网络结构和算法进行详细描述, 并在 IRIS 数据集上进行测试实验, 最后给出总结.

1 均衡学习矢量量化简介

PLVQ 是基于模糊学习规则(FLL)的一种广义学习矢量量化. 它是为解决学习矢量量化(LVQ)的问题而得出的, LVQ 是一种将大量非标注的矢量分类到某些给定类的聚类算法, 目前已经获得了较好结果, 但仍受到初始化问题和规范化问题的困扰.

基于 Hebb 的学习中, 假定 LVQ 网络的权学习规则在离散空间中具有以下规则:

$$\Delta m_i = \alpha_i u_i(x)(x - m_i) \quad (1)$$

式(1)中 x 表示输入矢量, m_i 表示神经元 i 的中心矢量, $u_i(x)$ 表示当给定 x 时, 神经元 i 的输出值, $\alpha_i = \alpha(1 - t/T)$ 指暂时学习率, 为了找出 $u_i(x)$ 的数学表达式和其物理意义, 考虑下面的丢失函数:

$$L = \sum_{i=1}^c \sum_{k=1}^N u_{ik}^q \sum_{j=1}^p (x_{kj} - m_{ji})^2 \quad (2)$$

式(2)中, u_{ik} 表示 x_k 与中心矢量 m_i 的匹配程度, 指数 q 控制模糊程度, 一般比 1 稍大, N 是数据的个数, c 是聚类的数目, 聚类数目等于输出神经元数目, p 为样本特征数, 即输入层节点数目.

使用最大模糊熵解释方法和规范化条件(对于每个样本 k 都有 $\sum_{j=1}^c u_{jk} = 1$), 得出以下最小化丢失函数的方案为

$$u_{ik} = \frac{\exp\left[-\frac{1}{2\sigma^2} \sum_{j=1}^p (x_{kj} - m_{ji})^2\right]}{\sum_{j=1}^c \exp\left[-\frac{1}{2\sigma^2} \sum_{j=1}^p (x_{kj} - m_{ji})^2\right]} \quad (3)$$

式(3)中, σ 是用户选取的非零值, 表示聚类中的模糊程度. σ 越小, 模糊度也越小, 一般选取

$$\sigma = 0.25\sqrt{p}$$

从等式(1)和式(3)可以看出每个输入以其对输出值的比值更新所有的权, 式(1)称为模糊学习规则(FLL)[3], 特别地, 当 $\sigma \rightarrow 0$, $u_{ik} = \{0, 1\}$ 时, 模糊学习规则演变为竞争学习规则(CLL)[2], 这种引入模糊隶属度的模糊学习规则的矢量量化算法为均衡学习矢量量化, 事实证明该均衡学习矢量

收稿日期:2008-05-26

基金项目:福建省自然科学基金(2006J0017)

作者简介:杨海燕(1980-)女, 湖南衡山人, 助教, 博士研究生. 研究方向:图像处理, 模式识别, 智能计算等.

量化避免了学习矢量量化的两个问题. 均衡学习矢量量化算法为:

(1) 参数设定: $2 \leq c \leq n, \sigma > 0, \epsilon > 0$ 和最大迭代次数 T

(2) 初始化 $\{m_i(0)\}$, 与学习速率 $\alpha_0 \in [0, 1]$

(3) for $t = 1, 2, \dots, T$

{for $k = 1, 2, \dots, N$

{用等式(3) 计算 $\{u_k\}$;

更新 $\{m_i(t)\} = m_i(t-1) + \alpha_0(1 - t/T)u_k(X_k - m_i);$
 $k+1\}$

计算 $E = \sum_{i=1}^c \sum_{j=1}^p |m_{ij}(t) - m_{ij}(t-1)|$

If $E < \epsilon$ 或 $t > T$ 终止; else $t = t+1\}$

2 面向特征选择的分类神经网络

该网络的结构可用图1描述, 它包括输入层、匹配层、检测层和输出层, 下面具体介绍该网络.

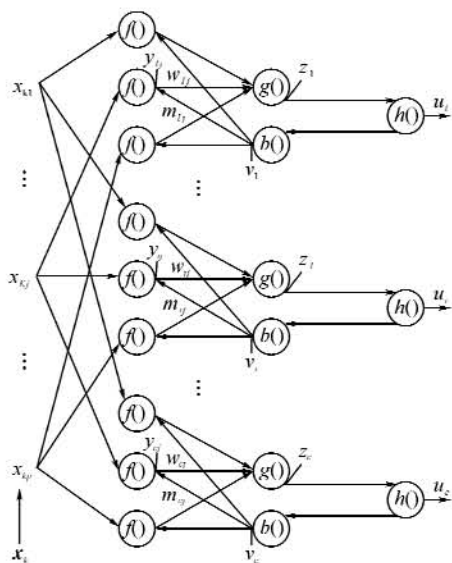


图1 面向特征选择的分类神经网络结构示意图

Fig.1 Structure profile of the classification neural network oriented to feature selection

2.1 网络结构

如图1所示, 该网络包含四层, 具体而言:

第一层: 输入层, 该层将有 p 个特征的样本 $x_k = (x_{k1}, \dots, x_{kj}, \dots, x_{kp})$ 输入网络.

第二层: 匹配层, 该层每个节点以两种方式接受输入, 一是正向输入层的各个特征值, 二是反馈回来的经过训练后的分类中心矢量, 该层结点的激励函数 $f(\cdot)$ 为比较函数, 而输入 y_{ij} 是两个输入值之差, 即:

$$y_{ij} = (x_{ij} - m_{ji}), j = 1, 2, \dots, p \quad (4)$$

第三层: 检测层, 该层有两种类型的节点对, 前向节点和反馈节点, 每个前向节点通过权连接

$\{w_{ij}\}$ 接收匹配层的 p 个节点的输入, 前向激励函数 $g(\cdot)$ 是高斯函数, 该前向结点的输出为

$$z_i = \exp\left[-\frac{1}{2\sigma^2} \sum_{j=1}^p w_{ij}^2 (x_{kj} - m_{ji})^2\right] \quad (5)$$

反馈节点通过固定与输出层节点的输入连接层的反馈路径, 反馈结点的激励函数 $b(\cdot)$ 是线性函数, 这保证每个神经元有相同的机会去匹配要到来的输入.

第四层: 输出层, 该层结点的个数为样本的分类个数 c , 输出节点的功能是对把所有检测节点的输出规范化, 决定每个输入结点各个分类的成绩, 也就是每个输入结点属于某类的隶属程度. 该层的输出节点有 c 个输入, 从检测层通过固定连接的 c 条路径得到. 此时, 神经元 i 的输出为

$$u_i = \frac{z_i}{\sum_{j=1}^c z_j} \quad (6)$$

2.2 学习规则

在该网络中, 对每个输入样本有两种类型的学习, 其中一个为中心矢量 $m_i = (m_{i1}, \dots, m_{ij}, \dots, m_{ip})$ 学习, 另一个是权 w_{ij} 学习, 中心矢量学习是非监督的, 更新规则基于 FLL, 即采用式(1)学习.

在权学习中, w_{ij} 表示样本特征 j 对于聚类 i 的贡献. 为了找出 w_{ij} 的更新规则, 引入误差函数:

$$E = \frac{1}{2} \sum_{k=1}^N \sum_{i=1}^c (u_i(x_k) - d_i)^2 \quad (7)$$

式(7)中 d_i 是输出层节点 i 的预期值, 即权学习是有监督的.

把等式(5)代入(6), 再带入(7), 采用基于微分方程的链式法则, 获得以下更新规则:

$$\Delta w_{ij} = \frac{\beta}{\alpha^2 s^2} (u^2(x_k) - d_i) \times \left(\sum_{j=1}^c z_j - z \right) w_{ij} z_i (x_{kj} - m_{ji})^2 \quad (8)$$

式(8)中, β 为学习率, w_{ij} 在 0 到 1 之间取值, $w_{ij} = 0$ 即特征 j 对聚类 i 无用, 为 1 表示 j 对聚类 i 的作用最大.

2.3 学习算法

该网络应用模糊隶属度函数进行学习, 初始化时采用随机选取的 c 个数据样本点作为聚类中心, 然后再利用神经网络的自学习功能, 逐步更新聚类中心矢量和权矢量. 具体步骤如下:

(1) $\alpha > 0, \alpha \in [0, 1], \beta > 0, \epsilon > 0$ 和最大迭代次数 T .

(2) 从 $\{x_k\} (k = 1, 2, \dots, N)$ 中随机产生 c 个样本初始化聚类中心 $\{m_i(0)\}$, 并对所有 i, j 置 $w_{ij}(0) = 1$.

(3)for $t = 1, 2, \dots, T$;
 {for $k = 1, 2, \dots, N$
 {用等式(3) 计算 $\{u_i\}$;
 计算 $\{m_i(l)\} = m_i(l-1) + \alpha_0(1 - t/T)u_{ik}(x_k - m_i)$;
 用等式(8) 更新 $\{w_{ij}(0)\}; k = k+1$ }
 使用等式(7) 计算 E ;
 If $E < \epsilon$ 或 $t > T$ 终止; else $t = t+1$ }

3 测试实验

为验证该网络的有效性,采用 IRIS 数据^[7]进行了实验. 实验数据有 150 个数据样本,每个样本用四个特征表示,可以分为三类,每类有 50 个样本,通常,已有的方法对其进行有监督分类的误差是 5 个,对非监督分类的误差是 15 个. 用 x_1, x_2, x_3, x_4 分别表示四个特征,具体为萼片长、萼片宽、花瓣长、花瓣宽,将该数据集用到提出的网络中,该网络的输入维数 p 为 4,输出层结点数 $c = 3$,选取运行参数 $\sigma = 0.5, \alpha = 0.01, \beta = 0.10, \epsilon = 2.0, T = 1000$,把 IRIS 数据集逐个输入,结果如表 1 实验一列的结果,可以看出,由于 w_{21}, w_{22} , 和 w_{31}, w_{32} 均为 0,可知特征 x_1 和 x_2 对聚类 c_2 和 c_3 没有贡献;而聚类 c_1 中特征 x_1 和 x_2 的作用可用特征 x_3 和 x_4 一起表现. 基于此,可以假设特征 x_1 和 x_2 无用,为证明假设正确,只用特征 x_3 和 x_4 再次进行实验,该次实验的运行参数与前次一样,结果如表 1 中的实验二所示,即误分类的个数与前次实验的结果一样,都是 5 个,而且权矢量为 $w_1 = (1.00, 1.00), w_2 = (1.00, 1.00), w_3 = (0.82, 0.97)$.

表 1 IRIS 数据实验结果

Table 1 The experiment result of IRIS data

	实验一(x_1, x_2, x_3, x_4)	实验二(x_3, x_4)
e	5	5
w_1	(1.00, 1.00, 0.95, 1.00)	(1.00, 1.00)
w_2	(0.00, 0.00, 1.00, 1.00)	(1.00, 1.00)
w_3	(0.00, 0.00, 0.82, 0.97)	(0.82, 0.97)

显然,用两个特征达到了与用四个特征一样的分类表现,所以就可以只选 x_3 (花瓣长)和 x_4 (花瓣宽)两个重要特征. 即在进行有监督分类的同时完成了特征选择目的.

4 结 语

以上针对模式识别中的特征选择与分类问题,提出了一种能够同时进行特征选择和有监督分类的分类神经网络. 该网络同时引入了两种类型的学习,即对中心矢量的学习和对权矢量的非监督学习,从 IRIS 数据进行实验结果得知:从原始四个特征中选取的两个特征与用所有特征进行分类的表现相当,达到了特征选择的目的,同时从分类误差和网络输出看,也完成了有监督分类的目的. 实验表明该分类神经网络具有分类和特征选择的能力.

参考文献:

- [1] 何 清. 模糊聚类分析理论与应用研究进展[A]. 模糊技术与神经网络选编[C]. 见刘曾良. 北京:北京航空航天大学出版社,1999:126-130.
- [2] Li R P. Proportional learning vector quantization[J]. Journal of Japan Society for Fuzzy Theory and Systems, 1998,10(6): 1129-1134.
- [3] Anthony M, John R. Fuzzy learning vector quantization for hyper-spectral coastal vegetation classification[J]. Remote Sensing of Environment, 2006,(100):512-530.
- [4] Wu Kuo-Lung, Yang Min-Shen. A fuzzy-soft learning vector quantization[J]. Neurocomputing, 2003,(55):681-697.
- [5] Mitra S, Hayashi Y. Neuro-fuzzy rule generation: survey in soft computing framework[J]. IEEE Trans Neural Networks, 2000,11(3):748-768.
- [6] 贺 玲,鲁汉榕. 一种改进的模糊学习矢量化神经网络[J]. 空军雷达学院学报,2001,15(1):33-35.
- [7] Anderson E. The IRISes of the Gaspé Peninsula[J]. Bull Amer IRIS Soc, 1939,(59):2-5.

A classification neural network oriented to feature selection

YANG Hai-yan^{1,2}, WEN Yi-ping³

(1. Department of Computer and Information Science, Fujian University of Technology, Fuzhou 350014, China;

2. College of Information Science and Engineering, Central South University, Changsha 410075, China;

3. School of Mathematics and Computer Science, Hunan University of Science and Technology, Xiangtan 411201, China)

Abstract: This paper proposed a new classification neural network. It can learn the center vector and

the connect weight simultaneously. The center vector is the center of classification, and the weight based on fuzzy membership represents the importance of feature. The features can be selected according to the final value of the weight, thus solving two major problems in pattern recognition: pattern classification and feature selection. The effectiveness of this method has been validated by IRIS data. The results show that the proposed network can select important features from the original features and maintain quite same performance as using the whole features.

Key words: feature selection; classification; center vector; connection weight

本文编辑:陈晓苹



(上接第113页)

参考文献:

- [1] 朱正德. 三坐标测量机的矢量检测功能及应用[J]. 测量, 2003, (1): 56-58.
- [2] 刘建国. 浅谈三坐标测量机对光滑工件尺寸的测量[J]. 测量, 2003, (9): 67-69.
- [3] 张洪娥, 黄劲东, 范文雷. 全最小二乘法及其在参数估计中的应用[J]. 自动化学报, 1995, (1): 40-47.
- [4] 黄宣国. 空间解析几何[M]. 上海: 复旦大学出版社, 2004: 24-53.

The application of orthogonal least square method in 3 dimensional coordinate measure

HU Xue-jun¹, WANG Jun-liang², WANG Gang²

(1. School of Electrical and Information Engineering, Wuhan Institute of Technology, Wuhan 430074, China;

2. College of Electronics Information, Wuhan University Wuhan 430072, China)

Abstract: This article describes theory of the orthogonal least square method. The space linear fit method is introduced via numerical example, which is on the condition of minimization of the square-sum of the point to line distance. The precision of the two methods are compared in 3 dimensional coordinate measure, which indicates that the orthogonal least square method is suitable for 3 dimensional coordinate measure machine.

Key words: orthogonal least square method; space linear; fit

本文编辑:陈晓苹