

基于信息扩散原理的 RBF 网络的分析与设计

王蕴红 刘国岁 王一丁

(南京理工大学电子工程系, 南京, 江苏 210094)

TP18

摘要 从信息扩散的角度对 RBF (Radial Basis Function) 网络进行了分析, 从理论上证明了 RBF 网络具有信息扩散功能, 并分析了其网络的物理意义, 说明了根据正态扩散的择近原则确定 RBF 网络中规划因子的合理性, 也说明了在训练样本数目大时用聚类方法确定中心参数的优越性. 实验结果证明这种方法设计的 RBF 网络分类器性能优于一般 RBF 网络分类器.

关键词 信息扩散, 正态扩散, 神经网络.

RBF网络

引言

近年来, 神经网络广泛应用于信号处理领域中, 由于各网络模型本身的特点, 在各个方面的作用也不尽相同, RBF 网络以其收敛快、物理意义明显等优点应用于模式识别中, 并取得了许多成果^[1~3]. 但 RBF 网络分类器的性能与它本身的结构及参数选择关系很大, 不适合的参数使得这种分类器效果很差. 如何合理选择 RBF 网络的参数已成为设计 RBF 网络分类器的关键问题. 以往选取网络隐层节点根据最小二乘算法^[3], 聚类算法^[4], 规划因子的选择更多地基于经验或其它主观性的方法. 训练样本不足也是使网络性能降低的主要原因, 如何充分利用已有信息, 获得较好的结果, 也是设计神经网络分类器的主要问题.

本文从信息扩散的角度分析了 RBF 网络的信息扩散功能, 证明了 RBF 网络能实现正态扩散, 由此根据正态扩散择近原则给出了确定规划因子的公式, 说明了在训练样本数目大时用聚类方法确定中心参数的合理性, 用这样设计的 RBF 网络作为分类器对三类战场目标进行识别, 并与一般 RBF 网络分类器的识别结果进行了比较.

1 信息扩散原理

在讨论信息扩散原理之前, 先引入完备样本空间及非完备样本空间概念^[5].

定义 1 设 w 是来自总体 O 的样本, 样本所可能取值的全体, 称为样本空间, 记为 W .

样本空间以样本表示出来, 当样本数目小时, 事实上就意味着样本不充分. 然而, 样本数目大也并不一定意味着知识充分. 事实上, 如总体分布是连续的, 无论样本的容量 m 如何大, 只要有限, 一旦两类相邻样本核间的距离取得太小, 它不能成为总体 O 的密度函数的较好逼近, 更不是精确的估计. 所以, 此时的大数目样本也并不充分, 即 W 对 O 来说是不完备

的. 下面我们再给出几个定义^[3]

定义 2 $\forall W \in W$, 设 γ 是对 W 进行某种数学处理的手段, 通过这种处理, 可由 W 作出对总体 O 的密度函数 $p(\nu)$ 的估计, 称 γ 为估计算子. 记所有估计的集合为 Γ . 设 $\gamma \in \Gamma$, 则

$$p_w^\gamma(\nu) = \gamma(W, \nu). \quad (1)$$

以 W 为由 γ 对 $p(\nu)$ 在 ν 点的估计值.

定义 3 设 $W \in W$, 如 $\forall \gamma \in \Gamma$, 使得对 $\forall \nu \in V$, 均有

$$|p(\nu) - p_w^\gamma(\nu)| > 0, \quad (2)$$

则称 W 为 O 的非完善样本, 简称 W 非完善.

定义 4 设 $W \in W$, 如 $\forall \gamma \in \Gamma$, 使得对 $\forall \nu \in V$, 均有

$$|p(\nu) - p_w^\gamma(\nu)| = 0, \quad (3)$$

则称 W 为 O 的完善样本, 简称 W 完善.

定义 5 $\forall W \in W$, 如 W 完善, 则称 W 为完备样本空间; 否则, 称为非完备样本空间.

1.1 非完备样本知识样本空间的模糊特性

从统计意义上来说, 非完备的 W 对总体的估计, 是近似估计. 如果增加 W 中的知识样本点, 使 W 趋于或达到完备. 从统计意义上讲, 由 W 对 O 的认识就趋于或达到清晰. W 从非完备到完备, 具有一种过渡趋势. 当 W 非完备时, 这种趋势表现在 W 的样本点上, 就是每个样本都具有发展成多个样点的趋势, 使得每个样点都充当“是周围未出现样点的代表”的角色. 而“周围”的边界是不清楚的、模糊的、富有弹性的, 所以每一个样点所提供的包括周围影响在内的信息总体是一个模糊信息. 显然, W 的样本点仍然可以是一些精确的观测数值, 本身并不模糊, 其模糊性是来自于 W 的非完备性.

当 W 非完备时, 其过渡性导致它具有模糊不确定性, 体现在每个样本点都有一定的影响域, 下面介绍的信息扩散原理就充分利用这种影响域信息.

1.2 信息扩散

定义 6 设 V 是样本 W 的基础论域, $\forall w_i \in W$, 设 w_i 的观测值为 ν_i , 则称 W 通过 w_i , 在 V 的 ν_i 点提供了一个量值为 1 的信息.

$\forall \nu \in V$, 由于其与 ν_i 点的关系所获得的信息称为是从 ν_i 点扩散来的信息, 而 ν_i 点的信息被分享的过程, 称为信息扩散过程.

定义 7 设 W 是知识样本, V 是基础论域, 所谓关于 W 的信息扩散, 就是 $W \times V$ 到 $[0, 1]$ 上的一个映射 $\mu: W \times V \rightarrow [0, 1]$.

且满足下列三个条件:

(1) $\forall w_i \in W$, 如 ν_i 是 w_i 的观测值, 则

$$\mu(w_i, \nu_i) = \sup \mu(w_i, \nu)$$

(2) $\forall w_i \in W$, $\mu(w_i, \nu)$ 随 $\|\nu_i - \nu\|$ 数值的增加而递减.

(3) $\forall w \in W$, $\int_V \mu(w, \nu) d\nu = 1$. 其中, 如 V 为离散情形, \int_V 表示成 \sum_V .

定义 8 $\forall w_i \in W$, 如 $\mu(w_i, \nu) = 1$, 称 μ 为平凡扩散, 否则, 称为非平凡扩散.

定义 9 设 $\mu(x)$ 为定义在 $(-\infty + \infty)$ 上的一个波雷尔可测函数, $\Delta_m > 0$ 为常数, 则称

$$f_m(\nu) = \frac{1}{m\Delta_m} \sum_{j=1}^m \mu\left(\frac{\nu - \nu_j}{\Delta_m}\right) \quad (4)$$

为总体密度 $p(\nu)$ 的一个扩散估计, 称 μ 为扩散函数, Δ_m 为窗宽. 这样, 解决信息扩散问题即是寻找一个适合的、满足定义 8 中关系的 $\mu\left(\frac{\nu - \nu_j}{\Delta_m}\right)$.

1.3 正态扩散方式

信息扩散原理: 设 $W = \{w_1, w_2, \dots, w_m\}$ 是知识样本, V 是基础论域, 记 w_i 的观测值为 ν_i , 设 $x = \varphi(\nu - \nu_i)$, 则 W 非完备时, 存在函数 $\mu(x)$, 使 ν_j 点获得的量值为 1 的信息可按 $\mu(x)$ 的量值扩散到 ν 中去, 且扩散所得的原始信息分布 $Q(\nu) = \sum_{j=1}^m \mu(x) = \sum_{j=1}^m \mu[\varphi(\nu - \nu_j)]$ 能更好地反映 W 内在规律.

用文献[5]数学物理方法推出了一种较实用的扩散方式——正态扩散方式, 这种方式应用简单, 又有一定精度, 其扩散函数为

$$\mu(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{x^2}{2\sigma^2}}, \quad (5)$$

上式满足定义 9 对扩散函数的一切要求, 设窗宽为 Δ_m , 则正态扩散估计为

$$f_m(\nu) = \frac{1}{m\Delta_m} \sum_{j=1}^m \left\{ \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{\left(\frac{\nu - \nu_j}{\Delta_m}\right)^2}{2\sigma^2}\right] \right\} = \frac{1}{\sqrt{2\pi m h}} \sum_{j=1}^m \exp\left[-\frac{(\nu - \nu_j)^2}{2h^2}\right], \quad (6)$$

式(6)中 $h = \sigma\Delta_m$ 为标准正态扩散 $\mu(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ 时的窗宽.

根据正态分布的两点择近原则, 可确定窗宽如下^[5]:

$$h = \begin{cases} 1.6987d/N \\ 1.4456d/N \\ 1.4230d/N \\ 1.4208d/N \end{cases} \begin{cases} 1 \leq N \leq 5, \\ 5 \leq N \leq 7, \\ 8 \leq N \leq 7, \\ 8 \leq N \leq 9, \\ 10 \leq N; \end{cases} \quad (7)$$

上式中 N 为样本容量, $d = \|a - b\|$, a, b 分别为样本观测到最大和最小状态.

2 RBF 网络的分析与设计

2.1 神经网络分类器研究信息扩散的必要性

神经网络分类器从某种意义上来说是通过学习产生对整个样本空间估计, 依此来实现分类功能. 神经网络对新样本的识别完全依赖于训练样本, 当训练样本数目不充分

或不足以反映整个样本空间的分布时,得到的关于样本空间的估计也是不准确的.这种样本空间的非完备性带来了样本的模糊性,为了充分学习这种“模糊性”,把每个样本的信息分给别的样本,即信息扩散是有必要的.RBF网络由于其本身结构的特点,可实现正态扩散从正态扩散的角度合理选择网络参数,具有实际意义.

2.2 基于信息扩散原理的RBF网络分析与参数选择

RBF网络的结构如图1所示.RBF网络是一个三层网络,在给定输入模式条件下,每个隐节点的输出为

$$y_k = \varphi(\|X - C_j\|), \quad (8)$$

隐节点函数即径向基函数可依需求选取,此处选高斯函数作为隐节点函数, C_j 表示高斯函数的中心,隐节点 j 的输出为

$$\begin{aligned} y_{kj} &= \varphi(\|X - C_j\|) \\ &= \exp\left[-\sum_{i=1}^L \frac{(x_i - c_{ij})^2}{2\sigma_{ij}^2}\right], \end{aligned} \quad (9)$$

σ_{ij} 为样本协方差矩阵中的元素,通常取样本协方差阵为对角阵,当对角线元素相等时记为 σ_j . Fig. 1 The structure of the RBF neural network 称为规划因子.

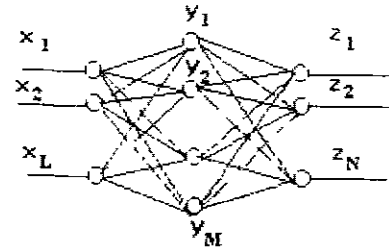


图1 RBF网络的结构

输出层单元的输出为隐层单元输出的线性加权和,第 i 个输出节点的输出表示为:

$$\begin{aligned} z_i &= \sum_h w_{ih} \varphi(\|X - C_j\|) + w_{i0} = \sum_{j=1}^m w_{ij} \exp\left[-\frac{(X - C_j)^2}{2\sigma_j^2}\right] + w_{i0} \\ &= \sum_{j=0}^m w_{ij} \exp\left[-\frac{(X - C_j)^2}{2\sigma_j^2}\right]. \end{aligned} \quad (10)$$

下面我们证明RBF网络可实现正态扩散功能.

证明:设隐层节点总数为 m ,将输入样本写成矢量形式,隐层第 j 节点输出为

$$y_{kj} = \exp\left[-\frac{(X - C_j)^2}{2\sigma_j^2}\right], \quad (11)$$

C_j 为中心因子,当样本数目较少时为所有的训练样本,当训练样本较多时为训练样本的一个子集,可与前面讨论的信息扩散中定义(6)的 ν_j 相对应, X 可与 ν 相对应,上式可写为

$$y_{kj} = \exp\left[-\frac{(\nu - \nu_j)^2}{2\sigma_j^2}\right], \quad (12)$$

σ_j 的选取有许多方法,此处取 $\sigma_j = h$, h 与式(6)中 h 的相同上式变为

$$y_{kj} = \exp\left[-\frac{(\nu - \nu_j)^2}{2h_j^2}\right], \quad (13)$$

输出层第 i 个节点输出为

$$z_i = \sum_{j=1}^m w_{ij} \exp\left[-\frac{(\nu - \nu_j)^2}{2h_j^2}\right], \quad (14)$$

而

$$\tilde{f}_m(\nu) = \frac{1}{\sqrt{2\pi mh}} \sum_{j=1}^m \exp\left[-\frac{(\nu - \nu_j)^2}{2h_j^2}\right],$$

当 $w_{ij}=1$ 时, $z_i = \sqrt{2\pi mh} \tilde{f}_m(\nu) + w_0$, RBF 网络实现了信息正态扩散功能.

当 $w_{ij}=1$ 时, z_i 与 $\tilde{f}_m(\nu)$ 也存在某种对应关系, 它们都是该函数 $\exp\left[-\frac{(\nu - \nu_j)^2}{2h_j^2}\right]$ 的加权和.

这种 RBF 网络的物理意义为: 在样本数目小的情况下, 样本空间是不完备的, 利用 RBF 网络的信息扩散功能可得出对总体样本概率密度估计, 利用别的网络模型得出的结果不如 RBF 网络好. 在前面的证明中, 实际上我们给出了 $\sigma_j = h$ 的假设, 故实现正态扩散功能的 RBF 网络, 规划因子的选择应依照正态扩散方式中选取窗 h 的原则选取, 可参见式(7). 在证明中, 设隐层节点数为 m , 当样本数目少时, 隐层节点数等于样本总数, 每个样本都是中心因子. 当样本数目大时, 隐层节点数取得过大, 会使网络过于庞大. 只能降低隐节点数目, 选取训练样本的子集作为中心因子, 也可把训练样本的一些聚类作为中心因子. 若取训练样本的聚类, 则该中心则近似为此分类的几何中点, 它们是基础论域用来吸收 W 提供信息的点, 这些点的集合构成 W 的离散基础论域 V , 对概率密度的估计是基于这些点的. 从估计的角度分析, 若随机选取某些样本作为中心, 基于这些点对整个样本空间的估计 $\tilde{f}_m(\nu)$ 必将是不精确的, 若中心因子是训练样本的某种聚类, 即为这些样本的代表, 则基于中心的估计 $\tilde{f}_m(\nu)$ 将近似于整个样本空间的分布.

从上面分析可知道, RBF 网络可直接用来作为信息扩散处理器, 它的信息扩散功能也可用来分析 RBF 网络作为分类器时的分类机理及性能评价.

3 实验及结论

用依照正态扩散方式设计的 RBF 网络以式(7)原则选取规划因子, 而一般 RBF 网络最简单选取规划因子的方法是取两个相邻样本点的欧氏距离, 我们分别用这种方式设计的 RBF 网络作为分类器, 对某雷达获得的三类战场目标——人、自行车、卡车进行识别, 以多卜勒谱作为这三类目标的特征, 将所得回波数据进行 1024 点 FFT, 取主要部分再经平均, 每个样本的特征维数为 20, 取 200 个典型样本, 一部分用于训练, 将全部样本都用来作识别样本. 样本数目小时为人、自行车、卡车各为 20 个, 样本数目非均匀时的情况为人样本 25 个, 自行车样本 10 个, 卡车样本 40 个. 实验结果见表 1.

上面两类网络中隐节点数等于样本总数, 中心因子即为每个训练样本, 但因为样本数目小或不均匀, 不足以产生对整个样本空间的精确估计, 从识别结果可看出, 基于扩散观点选择参数的 RBFN 比一般 RBFN 的识别率高.

表 1 样本数目小及样本非均匀的认识率

Table 1 The classification rates of deficient samples and uneven samples

样本类型	扩散 RBFN 的认识率 (%)		一般 RBFN 的认识率 (%)	
	小样本数目时	样本非均匀时	小样本数目时	样本非均匀时
人	76.7	73.8	60.2	52.8
自行车	76.4	64.6	71.7	36.4
卡车	87.0	88.2	74.6	81.2
平均认识率	80.0	75.6	68.6	56.8

当样本数目较大时,隐节点数取样本总数会使网络过于庞大,为减小隐节点数,我们采用两种方法设计 RBFN,第一种网络中心因子用模糊聚类的方法选取,聚类总数为隐节点数,规划因子的选取遵照式(7)规则,我们称这种网络为 FCRBFN,第二种网络中心因子随机选取,规划因子为两类相邻样本的欧氏距离一半,这种网络称为 RBFN,用这两种网络作为分类器,依然对前面三类战场目标进行识别,识别结果如表 2 所示。

表 2 FCRBFN 和 RBFN 的认识率

Table 2 The classification rates of FCRBFN and RBFN

网络类型 目标类型	FCRBFN 认识率 (%)		RBFN 认识率 (%)	
	12 个隐节点	20 个隐节点	12 个隐节点	20 个隐节点
人	84.2	80.2	68.0	76.4
自行车	82.1	82.4	56.2	60.4
卡车	86.8	87.6	74.7	72.0
平均认识率	84.8	83.4	66.3	69.6

从表 2 中可见,经过聚类得到中心因子的 FCRBFN 的认识率比一般 RBFN 的认识率有较大提高,模糊聚类选取的中心更加合理,按照信息扩散原理取规则因子使网络的效率提高,中心因子和规划因子的选取对网络的性能都有影响。

4 结语

本文从信息扩散的角度分析了 RBF 网络,从这方面讨论了提高 RBF 网络认识率的途径,理论和实验均说明了这种合理选择 RBF 网络参数设计 RBF 网络的方法是有效的。

本文旨在说明网络可实现信息扩散功能,实际上,利用扩散后的信息,再进行识别,认识率会提高,即以此 RBF 网络作为预处理器,将此结果再进行分类。

REFERENCES

- 1 Roth M W. *IEEE, Trans. NN*, 1990, 1: 28—43
- 2 Renals S, et al. *IJCNN*, 1990, 1: 461—467
- 3 Chen S, et al. *IEE, Proc. —F*, 1992, 139(6): 378—384
- 4 Vrokovnik G, et al. *IJCNN*, 1990, 1: 45—50
- 5 Huang C F, Wang J D. *Fuzzy Information Optimal Processing Technology and Applications*, Beijing: Press of Beijing University of Aeronautics and Astronautics (黄崇福, 王家鼎, 模糊信息优化处理技术, 北京:北京航空航天大学出版社), 1995, 24—64

ANALYSIS AND DESIGN OF RBF NEURAL NET BASED ON INFORMATION DIFFUSION

WANG Yun-Hong LIU Guo-Sui WANG Yi-Ding

(Department of Electronics Engineering, Nanjing University
of Science & Technology, Nanjing, Jiangsu 210094, China)

Abstract The RBF (Radial Basis Function) neural network was analyzed from the view point of information diffusion. The network's ability of information diffusion was proved in terms of theory. The practical meaning of the net was described. The paper illustrated the rational method of choosing the width by the normal diffusion principle and the advantage of choosing the center factor by clustering. The experimental results prove that the RBF net designed in this way has superior performances to the general RBF net.

Key words information diffusion, normal diffusion, neural network