



随机径向基函数神经网络的收敛性分析

张逾傲, 胡觉亮

(浙江理工大学理学院, 杭州 310018)

摘 要: 为了探究随机径向基函数神经网络的函数逼近能力, 运用随机权重前馈神经网络收敛性分析的方法对其进行收敛性分析。首先利用广义 δ 函数的性质构建一个被近似函数的极限积分表达式; 其次用蒙特卡罗方法计算这个表达式中的积分, 证明随机径向基函数神经网络可以逼近任意连续函数。同时, 从理论上分析了随机径向基函数神经网络的收敛特性, 发现其收敛误差随着隐藏层神经元节点的增加而逐渐减少, 表明其是一个高效的函数逼近器, 并且具有处理大数据问题的潜力。

关键词: 神经网络; 蒙特卡罗方法; 收敛性分析; 连续函数; 径向基函数; 随机径向基函数

中图分类号: TP391

文献标志码: A

文章编号: 1673-3851 (2019) 11-0835-06

Convergence analysis of stochastic radial basis function neural networks

ZHANG Yuao, HU Jueliang

(School of Sciences, Zhejiang Sci-Tech University, Hangzhou 310018, China)

Abstract: In order to explore the function approximation ability of radial basis function neural networks, convergence analysis method of random weight feedforward neural network was used to analyze its convergence. Firstly, the properties of generalized δ function were utilized to construct a limit integral expression of approximated functions, and then the integral in this expression was calculated by the Monte Carlo method to prove that the radial basis function neural networks can approximate any continuous functions. At the same time, the convergence characteristic of stochastic radial basis function neural network was theoretically analyzed, and it was found that the convergence error decreases gradually with the increase of neuron nodes at the hidden layer, which shows that it is an efficient function approximator and has the potential to deal with big data problems.

Key words: neural networks; Monte Carlo method; convergence analysis; continuous functions; radial basis function; stochastic radial basis function

0 引 言

人工神经网络自诞生以来, 理论和技术研究日益成熟, 应用范围也变得十分宽广。感知器作为最早的人工神经网络, 只能解决单一的线性问题, 而目前已创造出各种形式的人工神经网络, 并被广泛地运用到各个领域, 如语音识别^[1]、预测分析^[2]等。反

向传播算法 (Back propagation algorithm, BP)^[3] 由于具有自主学习参数的能力, 因而被广泛地运用于各种人工神经网络中, 但这种算法也存在局部极小值、收敛速度慢等问题^[4]。1992 年, Schmidt 等^[5]提出了随机权重的前馈神经网络 (Feedforward neural networks with random weights, FNNRW), 即隐藏层中的权重和偏置是按照某种方式随机选择的 (参

收稿日期: 2019-06-24 网络出版日期: 2019-10-08

基金项目: 国家自然科学基金项目 (11771393)

作者简介: 张逾傲 (1994-), 男, 浙江龙泉人, 硕士研究生, 主要从事神经网络方面的研究。

通信作者: 胡觉亮, E-mail: hujlh@163.com

数在 $[-1, 1]$ 之间均匀分布), 并利用最小二乘法对输出层的权重进行解析确定。但是这种随机设置隐藏层权重的方式并不是作为一种有效的算法提出的, 而只是一种探究前馈神经网络特性的简单方法。从近似理论的角度分析^[6], 隐藏层参数在 $[-1, 1]$ 之间均匀分布的方式并不能够作为一种随机模型去逼近任意连续函数(在概率为1的基础上)。因此, 关于它的近似能力和良好泛化性能的一些说法都是缺乏科学依据的^[7]。1995年, Igelnik等^[8]以一种构造性的方式指定隐藏层的权重和偏置称之为随机向量函数连接型网络(Random vector version of the functional-link net, RVFL), 并从理论上证明了这种网络的普遍逼近能力。RVFL网络与被近似函数之间的收敛误差依赖于隐藏层中神经元节点的数量, 节点数量越多收敛误差越小。同时, RVFL网络在处理大数据集问题方面具有巨大潜力, 因而受到广泛关注^[9-10]。由于只需要计算输出层权重这种良好的学习特性, 因此产生了许多关于如何计算RVFL网络输出层权重的算法^[11-12]。近几年, Wang等^[13-14]在随机权网络如何处理大数据集问题方面做了大量研究, 其成果凸显了这种神经网络的优势。此外, Scardapane等^[15]在2017年全面论述了有关这种神经网络的随机方法。由于这种随机方法对神经网络中的激活函数没有特定的约束, 所以本文希望可以将这种神经网络中一般的激活函数拓展到径向基函数。

另一方面, 具有相同平滑因子的径向基函数神经网络(Radial basis function neural networks, RBFNN)已经被Park等^[16]证明具有普遍的近似能力。因其具有分类能力强、学习收敛速率快、无局部极小值等良好特性, 被广泛地应用于各种领域。如何确定径向基函数神经网络的中心和平滑因子, 一直以来都是人们研究径向基函数神经网络的重点。关于径向基函数神经网络参数学习的主要算法, 其中一种是通过选取训练集样本作为径向基函数神经网络的中心, 再根据样本数据确定平滑因子。在这种学习算法中, 隐藏层节点数必须与样本数保持一致, 不利于网络结构的调整; 在面对大量的训练样本时, 也不利于输出层权重的计算。另一种是采用非监督学习和监督学习两个阶段的Moody算法^[17], 在非监督学习阶段用K-means聚类算法根据输入向量对训练样本进行聚类, 确定聚类中心和平滑因子。目前, 许多计算径向基神经网络中心的算法都从聚类角度出发, 但是聚类需要有一个确定样本之

间度量问题, 如何定义这种度量才能恰当地找到径向基网络的中心还需要研究。况且, 这种方法增加了一个非监督学习阶段, 同时也增加了时间成本, 降低了网络的性能。把随机设置隐藏层参数的方法与径向基函数神经网络的良好性能相结合, 能够有效地避免上述问题。随机径向基函数神经网络的中心和平滑因子采取随机选择的方式(在某个区间均匀分布), 所以不需要非监督学习阶段, 缩短了网络的训练时间; 而且其隐藏层的节点数可以任意增减, 不会影响网络结构。

本文运用Igelnik等^[8]随机权重前馈神经网络收敛性分析的思想, 通过蒙特卡罗计算积分的方法, 证明随机径向基函数神经网络(Radial basis function neural networks with random weights, RBFNNRW)同样能够逼近任意连续函数, 并且是一个有效的函数逼近器。

1 随机径向基函数神经网络

三层网络结构的径向基函数神经网络数学模型可以表示为:

$$f_{\sigma}(x) = \sum_i^n \beta_i g\left(\frac{\|x - y_i\|_2^2}{\sigma_i}\right) \quad (1)$$

其中: $x \in \mathbf{R}^m$ 是输入向量, $y_i = [y_{i1}, y_{i2}, y_{i3}, \dots, y_{im}]$ 表示隐藏层中第 i 个神经节点的中心; $\|x - y_i\|_2^2$ 表示输入向量与中心点之间欧氏距离的平方; g 是激活函数, 在径向基函数神经网络中通常为高斯函数; $\sigma_i \in \mathbf{R}$ 表示第 i 个神经节点激活函数的平滑因子, 它可以控制激活函数的宽度; β_i 是隐藏层到输出层的权重; n, m 分别表示隐藏层的节点数和输入向量的维度。本文中三层径向基函数神经网络的结构如图1所示。

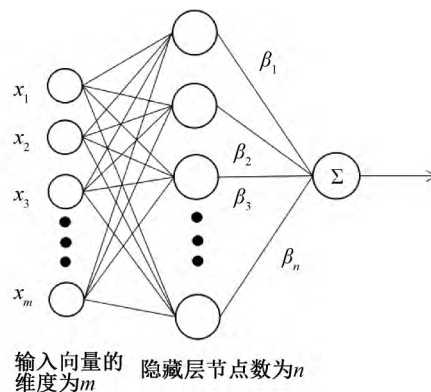


图1 径向基函数神经网络结构

随机径向基函数神经网络与一般的径向基神经网络的区别在于, 随机径向基函数神经网络中隐藏

层神经节点的中心 $y_i = [y_{i1}, y_{i2}, y_{i3}, \dots, y_{im}]$ 和平滑因子 $\sigma_i \in \mathbf{R}$ 以某种随机的方式生产(平均分布), 并且需要在输出层的权重 β_i 通过最小二乘法求出前确定。

2 随机径向基函数神经网络的收敛性分析

对任意连续函数 $f \in C(I^m)$ 和 RBFNNRW, 其中: $I^m = [0, 1]^m \subset \mathbf{R}^m$ 。并定义:

$$f_{\tilde{\omega}_n}(x) = \sum_i \beta_i g\{\theta_i \cdot [(x - y_i) \cdot (x - y_i)]\} \quad (2)$$

其中: $\tilde{\omega}_n = (n, \beta_1, \dots, \beta_n, y_1, \dots, y_n, \theta_1, \dots, \theta_n)$ 代表了这个网络的全部参数; $(x - y_i) \cdot (x - y_i)$ 表示两个向量的外积; \cdot 表示向量的内积。 $\tilde{\omega}_n$ 的随机部分定义为 $\lambda_n = (y_1, \dots, y_n, \theta_1, \dots, \theta_n)$, 假定 λ_n 被定义在概率测度分别为 μ, ∂ 的概率空间 $S_n(\Omega)$ 上, 假设 $S_n(\Omega), \mu, \partial$ 都依赖于确定的参数 Ω , 并且在学习阶段参数 Ω 需要被提前确定。概率测度 μ, ∂ 可以通过如下指定: $\theta_0 = (\theta_{01}, \theta_{02}, \dots, \theta_{0m})$ 均匀分布在 $V^m = [0, \Omega] \times [0, \Omega] \times \dots \times [0, \Omega]$ 中, 并且要求 $\theta_{01} = \theta_{02} = \dots = \theta_{0m}$, 同时 $y_0 = (y_{01}, y_{02}, \dots, y_{0m})$ 独立均匀分布在 I^m 中。所以 $(\theta_1, \theta_2, \dots, \theta_n)$ 和 (y_1, y_2, \dots, y_n) 分别表示来自 θ_0 和 y_0 分布的两个样本。

通过以上定义可以发现, 式(1)和式(2)只是在形式上不同。在以下证明中, 采用式(2)来分析随机径向基函数神经网络的收敛性。在任意紧集 $K(K \subset I^m)$ 中, 连续函数 f 与 $f_{\tilde{\omega}_n}(x)$ 之间的距离定义如下:

$$\rho_K(f, f_{\tilde{\omega}_n}) = \sqrt{\int_K [f(x) - f_{\tilde{\omega}_n}(x)]^2 dx} \quad (3)$$

定理1 对任意紧集 $K(K \subset I^m, K \neq I^m)$ 和绝对可积的激活函数 g , 有 $\int_{\mathbf{R}} g^2(x) dx < \infty$, 则存在一系列 RBFNNRW $\{f_{\tilde{\omega}_n}(x)\}$ 和一系列概率测度 $\{\mu, \partial\}$, 使得

$$\rho_K(f, f_{\tilde{\omega}_n}) \xrightarrow{n \rightarrow \infty} 0 \quad (4)$$

证明: 首先, 用多元函数积分的极限值来表示所要近似的目标函数。需要构造一个被积函数的核:

$$h_x(y, \theta) = g\{\theta \cdot [(x - y) \cdot (x - y)]\} \prod_{i=1}^m \sqrt{\theta_i} \quad (5)$$

并对激活函数 g 标准化, 使其满足 $\int_{\mathbf{R}^m} g(\sum_{i=1}^m z_i^2) dz = 1$, 其中 $z = \sqrt{\theta} \cdot (x - y)$ 。所以, 对任意的 $x, \theta \in \mathbf{R}^m$ 有 $\int_{\mathbf{R}^m} h_x(y, \theta) dy = 1$ 。当 $\theta_1, \theta_2, \dots, \theta_m \rightarrow \infty$, 函数

$h_x(y, \theta)$ 接近多维广义 δ 函数 $\delta_x(y)$ 。而广义 $\delta_x(y)$ 函数满足以下条件: 如果 $y \neq x$, 则 $\delta_x(y) = 0$ 并且有 $\int_{\mathbf{R}^m} \delta_x(y) dy = 1$ 。利用 $\delta_x(y)$ 函数的性质可以构造被近似函数的极限积分表达式:

$$f(x) = \lim_{\theta_1 \rightarrow \infty} \dots \lim_{\theta_m \rightarrow \infty} \int_{I^m} f(y) h_x(y, \theta) dy \quad (6)$$

下面证明在任意紧集 $K(K \subset I^m, K \neq I^m)$ 上, 式(6)成立。对任意 $\delta > 0$ 和任意的 $K \subset I^m, K \neq I^m$, 通过建立一个超立方体 $I_\delta^m = [\delta, 1 - \delta]^m$ 使得 $K \subset I_\delta^m \subset I^m, I_\delta^m \neq I^m$ 。利用 $\int_{\mathbf{R}^m} \delta_x(y) dy = 1$ 这个性质可以得到:

$$\begin{aligned} & \left| \int_{I^m} f(y) h_x(y, \theta) dy - f(x) \right| \\ & \leq \left| \int_{I^m} f(y) h_x(y, \theta) dy - \int_{I^m} f(x) h_x(y, \theta) dy \right| + \\ & \quad \int_{\mathbf{R}^m \setminus I^m} |f(x)| |h_x(y, \theta)| dy \\ & \leq \int_{I^m} |f(y) - f(x)| |h_x(y, \theta)| dy + \\ & \quad \int_{\mathbf{R}^m \setminus I^m} |f(x)| |h_x(y, \theta)| dy \quad (7) \end{aligned}$$

将变量 y 用 z 代替, 并定义:

$$\begin{aligned} V^m(x, \theta) &= [(x_1 - 1) \sqrt{\theta_1}, x_1 \sqrt{\theta_1}] \times \\ & \quad [(x_2 - 1) \sqrt{\theta_2}, x_2 \sqrt{\theta_2}] \times \dots \times [(x_m - 1) \\ & \quad \sqrt{\theta_m}, x_m \sqrt{\theta_m}], \\ I_\delta^m &= [(\delta - 1) \sqrt{\theta_1}, (1 - \delta) \sqrt{\theta_1}] \times [(\delta - 1) \sqrt{\theta_2}, \\ & \quad (1 - \delta) \sqrt{\theta_2}] \times \dots \times [(\delta - 1) \sqrt{\theta_m}, (1 - \delta) \sqrt{\theta_m}], \\ \sqrt{\theta}^{-1} &= (\sqrt{\theta_1}^{-1}, \sqrt{\theta_2}^{-1}, \dots, \sqrt{\theta_m}^{-1}), \end{aligned}$$

通过以上定义, 继续计算式(7):

$$\begin{aligned} & \int_{I^m} |f(y) - f(x)| |h_x(y, \theta)| dy \\ &= \int_{V^m(x, \theta)} |f(x - z \cdot \sqrt{\theta}^{-1}) - f(x)| |g(z^2)| dz \\ &\leq \int_{I_\delta^m(\theta)} |f(x - z \cdot \sqrt{\theta}^{-1}) - f(x)| |g(z^2)| dz, \end{aligned}$$

同时令 $S(\theta) = [-\sqrt[4]{\theta_1}, \sqrt[4]{\theta_1}] \times [-\sqrt[4]{\theta_2}, \sqrt[4]{\theta_2}] \times \dots \times [-\sqrt[4]{\theta_m}, \sqrt[4]{\theta_m}]$, 因此:

$$\begin{aligned} & \int_{I_\delta^m(\theta)} |f(x - z \cdot \sqrt{\theta}^{-1}) - f(x)| |g(z^2)| dz \\ &\leq \int_{S(\theta)} |f(x - z \cdot \sqrt{\theta}^{-1}) - f(x)| |g(z^2)| dz + \\ & \quad 2M \int_{\mathbf{R}^m \setminus S(\theta)} |g(z^2)| dz \\ &\leq \sup_{|T_i| \leq \frac{1}{\sqrt[4]{\theta_i}}, x \in I^m} |f(x - T) - f(x)| + \\ & \quad 2M \int_{\mathbf{R}^m \setminus S(\theta)} |g(z^2)| dz. \end{aligned}$$

其中: $M = \sup_{x \in I^m} f(x)$, $T = z \cdot \sqrt{\theta}^{-1}$, 当 θ 趋向于无穷的时候(可假设 $\theta \rightarrow \infty$, 即 $\theta_1 \rightarrow \infty, \dots, \theta_m \rightarrow \infty$), 最后一个不等式趋于 0。不等式中第一项趋于 0 是因为函数 f 在 I^m 中的连续性, 当 $\theta \rightarrow \infty$, T_i 的取值范围趋于 0, 所以 $f(x - T) - f(x) \rightarrow 0$ 。第二项趋于 0 是因为函数 $|g|$ 的可积性, 当 $\theta \rightarrow \infty$ 时, $S(\theta) \rightarrow \infty$, 所以第二项的积分域趋于 0, 最终可以得到这个积分趋于 0。另一方面,

$$\int_{\mathbb{R}^m \setminus I^m} |f(x)| |h_x(y, \theta)| dy \leq M \int_{\mathbb{R}^m \setminus \prod_{i=1}^m [-\delta\sqrt{\theta_i}, \delta\sqrt{\theta_i}]} |g(z^2)| dz \xrightarrow{\theta \rightarrow \infty} 0 \quad (8)$$

以上完成了第一步的证明。

其次, 考虑 $F(\theta) = \int_{I^m} f(y) h_x(y, \theta) dy$ 作为 θ 的函数, 并对这个函数应用 m 次洛必达法则, 可以得到:

$$\begin{aligned} & \lim_{\Omega_1 \rightarrow \infty} \lim_{\Omega_2 \rightarrow \infty} \cdots \lim_{\Omega_m \rightarrow \infty} \frac{1}{\prod_{i=1}^m \Omega_i} \int_{I^m \times \Omega^m} f(y) h_x(y, \theta) dy d\theta_{(m)} \\ &= \lim_{\Omega_1 \rightarrow \infty} \lim_{\Omega_2 \rightarrow \infty} \cdots \lim_{\Omega_m \rightarrow \infty} \frac{1}{\prod_{i=1}^{m-1} \Omega_i} \cdot \int_{I^m \times \Omega^{m-1}} f(y) \\ & \quad \prod_{i=1}^{m-1} \sqrt{\theta_i} \sqrt{\Omega_m} g\{(\theta_{m-1}, \Omega_m) \cdot [(x-y) \cdot (x-y)]\} \\ & \quad dy d\theta_{(m-1)} = \lim_{\Omega_1 \rightarrow \infty} \lim_{\Omega_2 \rightarrow \infty} \cdots \lim_{\Omega_m \rightarrow \infty} \int_{I^m} f(y) h_x(y, \Omega) dy = f(x). \end{aligned}$$

其中: $d\theta_{(m)} = \prod_{i=1}^m d\theta_i$; $\Omega^m = [0, \Omega_1] \times [0, \Omega_2] \times \cdots \times [0, \Omega_m]$ 。令 $\Omega_1 = \cdots = \Omega_m = \Omega$, 通过以上定义可得:

$$f(x) = \lim_{\Omega \rightarrow \infty} \frac{1}{\Omega^m} \int_{I^m \times V^m} f(y) g\{\theta \cdot [(x-y) \cdot (x-y)]\} \prod_{i=1}^m \sqrt{\theta_i} dy d\theta \quad (9)$$

其中: $V^m = [0, \Omega]^m$, 继续定义: $\tilde{\omega} = (y, \theta)$, $d\tilde{\omega} =$

$dy d\theta$, $W^m = I^m \times V^m$, $F_\Omega(y, \theta) = \frac{\prod_{i=1}^m \sqrt{\theta_i}}{\Omega^m} f(y)$, 所以式(9)可转变成:

$$f(x) = \lim_{\Omega \rightarrow \infty} \int_{W^m} F_\Omega(\tilde{\omega}) g\{\theta \cdot [(x-y) \cdot (x-y)]\} d\tilde{\omega} \quad (10)$$

式(10)代表函数 $f(x)$ 的极限积分表达式。

最后, 通过蒙特卡罗求积分的方法去计算式

(10)中等式右边的积分。因为 (y_1, y_2, \dots, y_n) 和 $(\theta_1, \theta_2, \dots, \theta_n)$ 是大小为 n 的样本, 分别均匀分布在 I^m, V^m 中, 所以 $(\tilde{\omega}_1, \tilde{\omega}_2, \dots, \tilde{\omega}_n)$ 是 W^m 中相对应的样本。通过以上步骤, 可以得到:

$$\int_{W^m} F_\Omega(\tilde{\omega}) g\{\theta \cdot [(x-y) \cdot (x-y)]\} d\tilde{\omega} \sim \frac{\Omega^m}{n} \sum_{i=1}^n F_\Omega(\tilde{\omega}_i) g\{\theta_i \cdot [(x-y_i) \cdot (x-y_i)]\} \quad (11)$$

根据蒙特卡罗计算积分的方法, 这个计算误差

为 $\frac{k \sqrt{\text{var} F_\Omega(\tilde{\omega}_i) g\{\theta_i \cdot [(x-y_i) \cdot (x-y_i)]\}}}{\sqrt{n}}$, 其中:

$\text{var} F_\Omega(\tilde{\omega}_i) g\{\theta_i \cdot [(x-y_i) \cdot (x-y_i)]\} = E\{\Omega^m F_\Omega(\tilde{\omega}_i) g\{\theta_i \cdot [(x-y_i) \cdot (x-y_i)]\} - \int_{W^m} F_\Omega(\tilde{\omega}) g\{\theta_i \cdot [(x-y_i) \cdot (x-y_i)]\} d\tilde{\omega}\}^2$ 。是随机变量 $F_\Omega(\tilde{\omega}_i) g\{\theta_i \cdot [(x-y_i) \cdot (x-y_i)]\}$ 的方差, E 表示数学期望, k 是依赖于可信度的常数。

所以, $\rho_K(f, f_{\tilde{\omega}_n})$ 可以通过如下计算得到:

$$\begin{aligned} & \int_K dx \left[\int_{W^m} F_\Omega(\tilde{\omega}) g\{\theta \cdot [(x-y) \cdot (x-y)]\} d\tilde{\omega} - \frac{\Omega^m}{n} \sum_{i=1}^n F_\Omega(\tilde{\omega}_i) g\{\theta_i \cdot [(x-y_i) \cdot (x-y_i)]\} \right]^2 = \\ & \frac{k^2}{n} \int_K dx E\{\Omega^m F_\Omega(\tilde{\omega}_i) g\{\theta_i \cdot [(x-y_i) \cdot (x-y_i)]\} - \int_{W^m} F_\Omega(\tilde{\omega}) g\{\theta \cdot [(x-y) \cdot (x-y)]\} d\tilde{\omega}\}^2 = \\ & \frac{b_{f,g,\Omega,m,k}}{n}. \text{ 当 } n \rightarrow \infty \text{ 时, } \frac{b_{f,g,\Omega,m,k}}{n} \rightarrow 0. \text{ 如果定义} \end{aligned}$$

$\beta_i = \frac{\Omega^m}{n} F_\Omega(\tilde{\omega}_i)$, $i=1, 2, \dots, n$ 。则式(11)可以重新改写成:

$$\int_{W^m} F_\Omega(\tilde{\omega}) g\{\theta \cdot [(x-y) \cdot (x-y)]\} d\tilde{\omega} \sim \sum_{i=1}^n \beta_i g\{\theta_i \cdot [(x-y_i) \cdot (x-y_i)]\} \quad (12)$$

结合式(10)和式(12)完成证明。

3 随机径向基神经网络是有效率的函数逼近器

通过对被近似函数增加一些限制, 假设被近似的函数满足李普希兹条件。即存在一个常数 $a > 0$, 使得对任意 $x, y \in I^m$, 有 $|f(x) - f(y)| \leq a \|x - y\|$ 。其中: $\|x - y\| = \sum_{i=1}^m |x_i - y_i|$, 这就把连续函数限制成满足上面条件的更光滑的函数。另外, 对激活函数也进行一些调优, 即让激活函

数 g 的定义域在区间 $\prod_{i=1}^m [-\alpha\sqrt{\theta_i}, \alpha\sqrt{\theta_i}]$ 上, 并记为 g_a 。可以有以下定理:

定理 2 对任意的 $f \in C(I^m)$ 并且满足李普希兹条件, 对任意的紧集 $K (K \subset I^m, K \neq I^m)$, 和激活函数 g_a 且满足定理 1 中激活函数的条件, 则存在一系列 RBFNNRW $\{f_{\tilde{\omega}_n}(x)\}$ 、一系列概率测度 $\{\mu, \partial\}$ 和常数 $b_{f,g,\Omega,m,k,a}$, 使得:

$$\rho_K(f, f_{\tilde{\omega}_n}) \leq \frac{b_{f,g,\Omega,m,k,a}}{\sqrt{n}} \quad (13)$$

证明: 通过前面的证明过程发现, RBFNNRW 逼近连续函数的思想主要分两步, 首先利用广义 δ 函数的性质构建一个被近似函数的极限积分表达式, 其次用蒙特卡罗方法来计算这个积分。所以, $\rho_K(f, f_{\tilde{\omega}_n})$ 可以通过式(14)来确定:

$$\rho_K(f, f_{\tilde{\omega}_n}) \leq \sup_{x \in I^m} |f(x) - f_{\Omega}(x)| + \rho_K(f_{\Omega}, f_{\tilde{\omega}_n}) \quad (14)$$

其中: $f_{\Omega}(x) = \int_{W^m} F_{\Omega}(\tilde{\omega}) g\{\theta \cdot [(x-y)(x-y)]\} d\tilde{\omega}$ 。通过取足够大的参数 Ω , 式(14)中第一项可以任意的小。但是在这种情况下, 决定第二项误差的方差会趋于无穷大,

$$\text{var}_{\Omega} = \int_K dx \frac{1}{|W^m|} \int_{W^m} \{ |W^m| F_{\Omega}(\tilde{\omega}) g\{\theta \cdot [(x-y)(x-y)]\} \}^2 d\tilde{\omega} - f^2(x) \xrightarrow{\Omega \rightarrow \infty} \infty。$$

如果能够保证 $\sup_{x \in I^m} |f(x) - f_{\Omega}(x)|$ 趋于 0 的速率比方差 var_{Ω} 趋于无穷的速率要快, 那么得到的误差 $\rho_K(f, f_{\tilde{\omega}_n}) \leq \frac{b_{f,g,\Omega,m,k,a}}{\sqrt{n}}$ 。但是这种情况对函数 f 的限制严重, 所以更改激活函数 g 的定义域为 $\alpha(n)\Omega$, 其中 $\alpha(n) \xrightarrow{n \rightarrow \infty} 0$, 这样可以通过固定 Ω 的值来保证方差 var_{Ω} 有界, 同时有

$$\sup_{x \in I^m} |f(x) - f_{\Omega}(x)| \xrightarrow{n \rightarrow \infty} 0 \quad (15)$$

通过证明发现, 采用这种方式同样可以得到 $\rho_K(f, f_{\tilde{\omega}_n}) \leq \frac{b_{f,g,\Omega,m,k,a}}{\sqrt{n}}$ 。下面用一个引理来证明式(13)成立。

引理 1 g_a 表示在区间 $\prod_{i=1}^m [-\alpha\sqrt{\theta_i}, \alpha\sqrt{\theta_i}]$ 上的激活函数, 其中 $\alpha, \theta_1, \theta_2, \dots, \theta_m$ 是任意的正整数。重新定义核函数:

$$h_{x,a}(y, \theta) = g_a\{\theta_i \cdot [(x-y_i)(x-y_i)]\} \prod_{i=1}^m \sqrt{\theta_i},$$

使得在任意紧集 $K (K \subset I^m, K \neq I^m)$ 上:

$$f(x) = \lim_{a \rightarrow 0} \int_{I^m} f(y) h_{x,a}(y, \theta) dy。$$

事实上,

$$\begin{aligned} & \left| \int_{I^m} f(y) h_{x,a}(y, \theta) dy - f(x) \right| \\ & \leq \int_{\sup g_a} |f(x - z \cdot \sqrt{\theta}^{-1}) - f(x)| |g(z^2)| dz \\ & \leq \sup_{z \in \sup g_a} |f(x - z \cdot \sqrt{\theta}^{-1}) - f(x)| \\ & \leq k \sup_{z \in \sup g_a} |z \cdot \sqrt{\theta}^{-1}| \leq k\alpha。 \end{aligned}$$

其中: $\sup g_a$ 表示函数 g_a 的定义域, 当 n 趋于无穷的时候 α 趋向 0。因此可以通过有界的 θ , 用 $\int_{I^m} f(y) h_{x,a}(y, \theta) dy$ 近似函数 $f(x)$, 再通过有界的 Ω , 用 $\int_{W^m} F_{\Omega}(\tilde{\omega}) g\{\theta \cdot [(x-y)(x-y)]\} d\tilde{\omega}$ 逼近函数 $f(x)$, 所以 $b_{f,g,\Omega,m,k,a}$ 为一个有界的常数。

本文运用了随机权重前馈神经网络收敛性证明的思想方法, 因为在证明过程中构造了不同的核函数, 所以可以有效地避免文献[8]中一个错误的证明步骤, 这个错误的证明步骤已被文献[18]指出。通过上文的证明过程可知, 随机径向基函数神经网络与被近似函数之间的收敛误差可以由表达式 $\frac{b_{f,g,\Omega,m,k,a}}{\sqrt{n}}$ 决定。由于 $b_{f,g,\Omega,m,k,a}$ 是一个有界的常数,

所以与收敛误差有关的变量就只有隐藏层中神经元的节点数 n 。随着神经元节点数地增加, 收敛误差逐渐降低趋于 0, 很容易就能达到人们所能接受的误差范围, 因此该神经网络是一个有效率的函数逼近器。这种随机设定径向基函数中心和平滑因子的方式, 使得在神经网络中只需要计算输出层权重这一种参数即可, 并不需要非监督学习阶段, 因而提高了网络的学习效率。并且, 采用最小二平方法求输出层权重是一种线性运算, 在计算上十分迅速。输入向量维度的增加也并不影响随机径向基函数的收敛误差, 因此该神经网络在处理大数据集问题上具有巨大的潜力。

4 结 论

随机径向基函数神经网络中部分参数的随机选择方式, 是对现有确定径向基函数神经网络中心和平滑因子方式的一种重要补充。本文证明了随机径向基函数神经网络不仅能够逼近任意连续函数, 而且还是一个有效率的函数逼近器。此外, 相较于其他径向基函数神经网络, 随机径向基函数神经网络中部分参数可采用随机选择方式确定, 且输出层权

重可以通过线性计算求解。这些特性使得随机径向基函数神经网络能够缩短网络的训练时间,并且在处理大数据集任务上具有巨大潜力。

参考文献:

- [1] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep Convolutional neural networks [C]// Proceedings of the 25th International Conference on Neural Information Processing Systems. New York: Curran Associates, 2012: 1097-1105.
- [2] Salkuti S R. Short-term electrical load forecasting using radial basis function neural networks considering weather factors[J]. Electrical Engineering, 2018, 100 (3): 1985-1995.
- [3] Werbos P J. Generalization of Backpropagation with application to a recurrent gas market model[J]. Neural Networks, 1988, 1(4):339-356.
- [4] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016: 97-117.
- [5] Schmidt W F, Kraaijveld M A, Duin R P W. Feedforward neural networks with random weights [C]//11th IAPR International Conference on Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 1992:1-4.
- [6] Li M, Wang D H. Insights into randomized algorithms for neural networks: Practical issues and common pitfalls [J]. Information Sciences, 2017, 382/383: 170-178.
- [7] Wang D H, Li M. Stochastic Configuration Networks: fundamentals and algorithms[J]. IEEE Transactions on Cybernetics, 2017, 47(10):3466-3479.
- [8] Igel'nik B, Pao Y H. Stochastic choice of basis functions in adaptive function approximation and the functional-link net[J]. IEEE Transactions on Neural Networks, 1995, 6(6):1320-1329.
- [9] Tyukin I Y, Prokhorov D V. Feasibility of random basis function approximators for modeling and control[C]// 2009 IEEE International Conference on Control Applications. IEEE, 2009:1391-1396.
- [10] Ye H L, Cao F L, Wang D H, et al. Building feedforward neural networks with random weights for large scale datasets [J]. Expert Systems with Applications, 2018, 106:233-243.
- [11] Cao F L, Ye H L, Wang D H. A probabilistic learning algorithm for robust modeling using neural networks with random weights[J]. Information Sciences, 2015, 313:62-78.
- [12] Cao F L, Wang D H, Zhu H Y, et al. An iterative learning algorithm for feedforward neural networks with Random Weights[J]. Information Sciences, 2016, 328:546-557.
- [13] Wang D H, Cui C H. Stochastic configuration networks ensemble with heterogeneous features for large-scale data analytics[J]. Information Sciences, 2017, 417:55-71.
- [14] Wang D H, Li M. Stochastic configuration networks: Fundamentals and algorithms[J]. IEEE Transactions on Cybernetics, 2017, 47(10):3466-3479.
- [15] Scardapane S, Wang D H. Randomness in neural networks: An overview [J]. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2017, 7(2):1200-1218.
- [16] Park J, Sandberg I W. Universal approximation using Radial-Basis-Function networks [J]. Neural Computation, 1991, 3(2):246-257.
- [17] Moody J, Darken C J. Fast learning in networks of locally-tuned processing units [J]. Neural Computation, 1989, 1(2):281-294.
- [18] Li J Y, Chow T W S. Comments on "Stochastic choice of basis functions in adaptive function approximation and the functional-link net" [J]. IEEE Transactions on Neural Networks, 1997, 8(2):452-454.

(责任编辑:康 锋)