

二阶 Newton 法训练径向基函数神经网络的算法研究

蔡 珣¹ 陈 智¹ Kanishka Tyagi² 于 宽³ 李子强¹ 朱 波⁴

¹(山东大学计算机科学与技术学院 济南 250101)

²(德克萨斯大学阿灵顿分校电子工程系 美国阿灵顿 76019)

³(山东建筑大学材料科学与工程学院 济南 250101)

⁴(山东大学材料科学与工程学院 济南 250061)

(caixunzh@sdu.edu.cn)

Second Order Newton's Method for Training Radial Basis Function Neural Networks

Cai Xun¹, Chen Zhi¹, Kanishka Tyagi², Yu Kuan³, Li Ziqiang¹, and Zhu Bo⁴

¹(School of Computer Science and Technology, Shandong University, Jinan 250101)

²(Department of Electrical Engineering, University of Texas at Arlington, Arlington, USA 76019)

³(School of Material Science and Engineering, Shandong Jianzhu University, Jinan 250101)

⁴(School of Materials Science and Engineering, Shandong University, Jinan 250061)

Abstract A hybrid two-step second-order batch approach is presented for constructing and training radial basis function (RBF) neural networks. Unlike other RBF neural network learning algorithms, the proposed paradigm uses Newton's method to train each set of network parameters, i. e. spread parameters, mean vector parameters and weighted distance measure (DM) coefficients and output weights parameters. For efficiently calculating the second-order equations of Newton's method, all the optimal parameters are found out using orthogonal least squares (OLS) with the multiply optimal learning factors (MOLFs) for training mean vector parameters. The simulation results of the proposed hybrid training algorithm on a real dataset are compared with those of the recursive least square based RBF (RLS-RBF) and Levenberg-Marquardt method based RBF (LM-RBF) training algorithms. Also, the analysis of the training performance for optimization of each set of parameters has been presented. The experimental results show that the proposed hybrid optimal weighted DM training algorithm, which is based on the optimization of the mean vectors, weighted DM coefficients and spread parameters, has significant improvement on training convergence speed compared with that of RLS-RBF and has very less computation cost compared with that of LM-RBF. It confirms that Newton's method solved by OLS is a significantly valuable method for training the RBF neural network.

Key words radial basis function (RBF) neural network; Hessian matrix; Newton's method; orthogonal least squares (OLS); weighted distance measure (weighted DM); multiply optimal learning factors (MOLFs)

摘 要 提出了一种混合加权距离测量 (weighted distance measure, weighted DM) 参数的构建和训练 RBF (radial basis function) 神经网络的两步批处理算法. 该算法在引进了 DM 系数参数的基础上, 采用 Newton 法分别对径向基函数的覆盖参数、均值向量参数、加权距离测度系数以及输出权值进行了优化,

并在优化过程中利用 OLS(orthogonal least squares)法来求解 Newton 法的方程组. 通过实验数据, 不仅分析了 Newton 法优化的各个参数向量对 RBF 网络训练的影响, 而且比较了混合优化加权 DM 与 RLS-RBF(recursive least square RBF neural network)网络训练算法的收敛性和计算成本. 所得到的结论表明整合了优化参数的加权 DM-RBF 网络训练算法收敛速度比 RLS-RBF 网络训练算法更快, 而且具有比 LM-RBF(Levenberg-Marquardt RBF)训练算法更小的计算成本, 从而说明 OLS 求解的 Newton 法对优化 RBF 网络参数具有重要应用价值.

关键词 径向基函数神经网络; Hessian 矩阵; Newton 法; 正交最小二乘法; 网络参数优化; 最优学习因子

中图法分类号 TP183

径向基函数(radial basis function, RBF)网络是一种有监督 3 层前向网络, 用于插补、概率密度函数估计和平滑多元函数拟合等^[1-2]数学计算. RBF 网络最先是为了解决实数多元插补问题而被引入. 给定足够数量的径向基函数, 传统的 RBF 网络训练算法能够在一个紧致域内很好地拟合任意的多元连续函数^[3]. RBF 网络现已广泛应用于多个领域, 如初期缺陷检测^[4]、粒子群优化算法^[5]、乳腺癌诊断^[6]、涡流发电机的建模^[7]、实时潮水位预测^[8]和时序周期性预测^[9]等. 同时, RBF 网络在机械加工、网络拥塞预测方面也有着广泛的应用^[10-11].

构建 RBF 网络可分成三大类型:

1) 无训练型. 该类型网络的所有参数都是预先固定, 不需要训练. 这种情况下, 由于网络规模太大而且速度太慢, 因此该类型 RBF 网络没有多少应用价值.

2) 半训练型. 该类型网络的隐含层参数是通过启发式或一些聚类算法预先计算得到, 且保持不变. 训练只针对输出层参数采用最小二乘法进行.

3) 完全训练型. 该类型网络所有的参数和权值都通过一种训练算法而不断更新学习. 例如, 文献[12]提出了一个混合学习算法, 它整合聚类算法如 *k*-means 算法来确定 RBF 网络均值向量以及监督学习来更新隐含层到输出层的输出权值; 文献[13]提出了梯度训练算法来更新 RBF 网络参数(均值向量, 覆盖参数向量)以及输出权值.

无论半训练型还是完全训练型构建的 RBF 网络, 采用的训练算法主要是一阶和二阶学习方法^[14], 其中一阶方法最为普及. 该类算法能够平衡收敛性能要求和训练速度要求, 如广泛采用的梯度下降(*gradient descent*)学习法. 尽管一阶方法比较适合大规模问题^[13], 但存在对输入均值和增益因子敏感性强的问题, 因而难以最优化. 与 Newton 法有关的一系列算法均属于二阶法^[15], 如 Levenberg-Marquardt(LM)法^[16]. 相比一阶方法, 二阶方法具

有更高的精确度, 但由于计算复杂度高, 因而更适合于小规模问题. 因此, 整合最快梯度下降(*steepest descent*)一阶方法和二阶 Newton 法的方法成为解决非约束优化问题的更有前景^[17-19]的算法.

在训练模式方面, 除了少部分 RBF 网络采用串行训练模式外, 如通用生长剪枝 RBF(*generalized growing and pruning RBF*, GGAP-RBF)网络^[20], 大多数 RBF 网络学习算法采用批处理模式. 为了缩减网络规模, 有些研究者提出通过优化隐含单元数目的方法获得比传统大型 RBF 网络更紧致的网络^[21-23]. 然而, 为了选择一个子集网络, 这类算法需要在全连接网络中遍历所有数据, 再采用一种基于正交最小二乘(*orthogonal least squares*, OLS)或规格化 OLS 对全连接网络进行子集选择方法, 因此算法复杂度较大. 为了减低算法复杂度并获得更好的泛化能力, 研究者提出了一种整合了前向子集选择和零阶正则化方法的正则化前向子集选择(*regularized subset selection*, RFS)法^[24].

国内对 RBF 网络训练算法的研究也有很多, 如王学雷等人^[25]研究了 RBF 网络的在线学习算法, 谭建辉^[26]研究了 RBF 网络的再学习算法, 并将 RBF 网络的再学习问题巧妙地转化为矩阵求逆的附加运算, 从而取得了良好效果. 尽管如此, RBF 网络的收敛速度和泛化性仍是一个有待解决的问题.

针对上述问题, 本文余下部分将首先回顾传统 RBF 网络的结构及相关的数学描述, 然后对传统训练算法进行分析, 并在此基础上提出了一组基于 Newton 法的 RBF 网络二阶训练算法, 并通过大量实验验证了这组算法的有效性.

1 RBF 网络结构回顾

本节我们将首先介绍 RBF 网络的一些描述符号和结构, 然后讨论现有 RBF 训练算法存在的一些问题.

1.1 传统 RBF 结构和标记

不失一般性,我们选定传统 3 层全连接 RBF 网络结构作为研究对象,如图 1 所示:

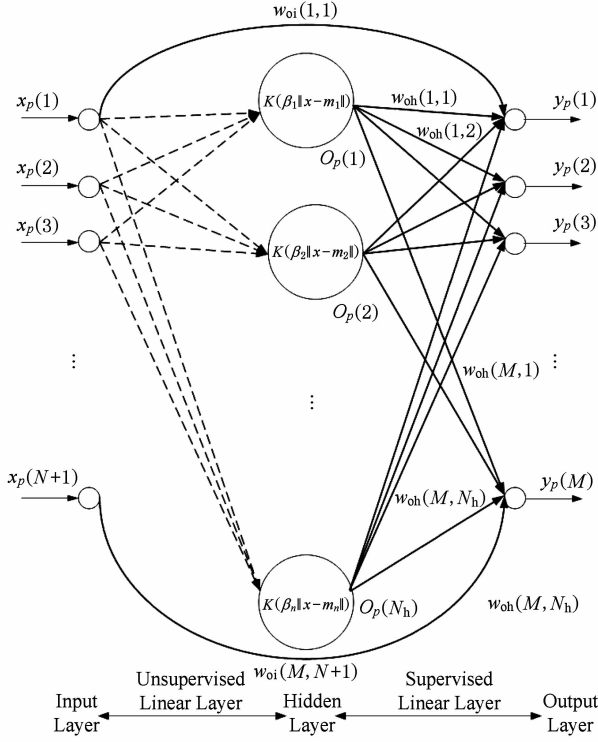


Fig. 1 Topology of a fully connected RBF neural network.

图 1 全连接 RBF 神经网络的拓扑结构

假定训练数据集包含 N_{vp} 个训练样本集 $\{\mathbf{x}_p, \mathbf{t}_p | 1 \leq p \leq N_{vp}\}$, 其中 $\mathbf{x}_p$ 表示第 p 个样本的输入向量, $\mathbf{t}_p$ 表示对应的输出向量, 它们的维度分别是 N 和 M . 通常为了计算方便, 输入层到隐含层的偏置向量视为第 $N+1$ 个输入单元到隐含层的权值向量并且 $x_p(N+1)=1$, 因此输入向量成为 $N+1$ 维向量, 即 $\mathbf{x}_p = [x_p(1), x_p(2), \dots, x_p(N+1)]^T$.

如图 1 所示, 各个输入单元直接与有 N_h 个结点的隐含层相连. 第 k ($k=1, 2, \dots, N_h$) 个隐含层单元对应输入向量的第 k 簇均值向量 $\mathbf{m}_k$, 也称为核向量或中心向量. 对于第 p 个训练样本, RBF 网络的第 k 个隐含层单元的激活函数的输入将不是传统人工神经网络的加权求和函数, 而是输入 $\mathbf{x}_p$ 与第 k 个隐含单元相关的均值向量 $\mathbf{m}_k$ 的相似度函数, 即

$$net_p(k) = \sum_{n=1}^N (x_p(n) - m_k(n))^2, \quad (1)$$

其中, $m_k(n)$ 表示第 k 个隐含层单元对应的第 n 个输入单元的簇均值.

第 p 个样本的第 k 个隐含单元的激活函数的输出 $O_p(k)$ 采用高斯基函数计算得到, 即

$$O_p(k) = e^{-\beta(k)net_p(k)}, \quad (2)$$

其中, $\beta(k)$ 称为第 k 个隐含单元的覆盖参数值. 文献 [15] 将 $\beta(k)$ 定义为第 k 个隐含单元以 $\mathbf{m}_k$ 为均值向量的高斯函数宽度 σ 的倒数, 并提出了启发式算法对所有高斯函数分配固定宽度, 即 $\sigma = d_{\max} / \sqrt{2N_h}$, 其中 $d_{\max}$ 为选定聚类中心间的最大距离.

为了保证第 k 个隐含单元可激活, 其输出应当满足以下条件:

$$O_p(k) \geq e^{-1}. \quad (3)$$

在输出层中, 令 $w_{oh}(i, k)$ 表示第 k 个隐含单元到第 i 个输出 $y_p(i)$ 的权值, 则第 p 个训练样本的输出 $y_p(i)$ 为

$$y_p(i) = \sum_{k=1}^{N_h} w_{oh}(i, k) O_p(k) + \sum_{n=1}^N w_{oi}(i, n) x_p(n) + b_i, \quad (4)$$

其中, $w_{oi}(i, n)$ 为旁路权值, 即将第 n 个输入单元和第 i 个输出单元直接相连, 而 b_i 为第 i 个输出单元的偏置值. 式(4)中输出可以写成向量形式为

$$\mathbf{y}_p = \mathbf{W}_{oi} \cdot \mathbf{x}_p + \mathbf{W}_{oh} \cdot \mathbf{O}_p + \mathbf{b}. \quad (5)$$

输出层的输入可以整合为

$$\bar{x}_p(n) = \begin{cases} x_p(n), & n = 1, 2, \dots, N, \\ 1, & n = N+1, \\ O_p(n-N-1), & n = N+2, \dots, N_{un}, \end{cases} \quad (6)$$

其中, $N_{un} = N+1+N_h$. 整合后的输入也可以用向量形式表示为

$$\bar{\mathbf{x}}_p = [\mathbf{x}_p^T; 1; \mathbf{O}_p^T]. \quad (7)$$

为方便表达, 输出权值可以整合为

$$w_o(i, k) = \begin{cases} w_{oh}(i, k), & 1 \leq k \leq N, \\ w_{oi}(i, N+1), & k = N+1, \\ w_{oi}(i, k-N-1), & N+2 \leq k \leq N_{un}, \end{cases} \quad (8)$$

这里 $w_{oi}(i, N+1)$ 为增补输入单元 $x_p[N+1]$ 到第 i 个输出单元的旁路权值, 实际上等价于式(4)中的 b_i .

如果令 $\mathbf{W}_o$ 代表输出权值矩阵, 则式(8)可以用向量形式表示为

$$\mathbf{W}_o = [\mathbf{W}_{oi}; \mathbf{W}_{oh}]. \quad (9)$$

结合式(7)和式(9), 式(4)可简写为

$$\mathbf{y}_p = \mathbf{W}_o \cdot \bar{\mathbf{x}}_p, \quad (10)$$

或表示为

$$y_p(i) = \sum_{k=1}^{N_{un}} w_o(i,k) \bar{x}_p(k). \tag{11}$$

如果将第 p 个样本的训练误差用方差表示为

$$E_p = \sum_{i=1}^M [t_p(i) - y_p(i)]^2, \tag{12}$$

则在批处理模式中,网络的整体误差函数将采用均方差(mean square errors, MSE)表示,即

$$E = \frac{1}{N_{vp}} \sum_{p=1}^{N_{vp}} E_p = \frac{1}{N_{vp}} \sum_{p=1}^{N_{vp}} \sum_{i=1}^M [t_p(i) - y_p(i)]^2 = \frac{1}{N_{vp}} \sum_{p=1}^{N_{vp}} (t_p - y_p)^T (t_p - y_p), \tag{13}$$

其中 t_p 和 y_p 为列向量. 权值优化通过最小化 MSE 实现^[27].

为了 RBF 网络的均值向量、覆盖参数向量以及训练输出权值 $\mathbf{W}$. 等各个网络参数的训练,一种传统的 RBF 训练算法是 RLS-RBF (recursive least squares RBF)^[28] 算法. 该算法首先采用 k -mean 初始化隐含层参数,即均值向量、覆盖参数向量等,并固定下来,然后利用梯度下降法来训练输出层权值.

1.2 现有 RBF 训练算法的主要问题

现有 RBF 训练算法受限于 5 个主要问题:

- 1) 网络参数没有得到很好的训练.
- 2) 有噪声的输入干扰网络参数正确性.
- 3) 采用聚类获得的均值向量等网络参数没有进行优化.
- 4) 输出权值采用梯度下降法或递归法进行训练,存在误差振荡易陷入局部最优的问题.
- 5) Newton 法训练算法当其 Hessian 矩阵为奇异矩阵时,不能解出所有权值.

2 本文工作

针对现有算法存在的问题,本文提出了一种混合 RBF 网络参数训练的算法,内容包括:对各类网络参数初始化进行改进;为优化距离测量(distance measure, DM)引进加权系数向量 $\mathbf{c}$ 并对其进行优化;引进均值向量学习因子向量以及覆盖参数向量 β 的学习因子向量并进行了优化;采用输出权值优化(output weights optimization, OWO)法对输入权值进行优化;采用 OLS 法求解 Newton 法线性方程组,避免 Hessian 矩阵的奇异性.

2.1 均值向量及加权距离测量系数向量初始化

在本文中,均值向量 $\mathbf{m}_k$ 的初始化通过采用 SOM (self-organizing map) 对输入向量进行聚类得到.

传统欧几里得 DM 训练算法是一种标准聚类训练算法,然而很多文献,特别是在语言处理领域^[29],认为使用加权的 DM 相比传统欧式距离可以获得质量更好的结果和更少的谱失真. 因此,本文引进了加权距离测量(distance measure, DM)系数向量 $\mathbf{c}$ 并作为后续参数训练的评测. 采用加权 DM 后,隐含层第 k 个单元的输入函数改写为

$$net_p(k) = \sum_{n=1}^{N+1} c(n) (x_p(n) - m_k(n))^2. \tag{14}$$

令 $\sigma(n)$ 表示所有输入向量 $\mathbf{x}_p$ 的第 n 个单元的标准差,则 $c(n)$ 可以初始化为

$$c(n) = \frac{\sigma^2(n)}{\sum_{m=1}^N \frac{1}{\sigma^2(m)}}. \tag{15}$$

2.2 覆盖参数向量 β 的初始化

为了初始化覆盖参数向量 β ,所提出的算法引入一个 $N_h \times N_h$ 加权均值向量距离矩阵 $\mathbf{D}_m$,元素 $d_m(m_k, m_u)$ 表示为

$$d_m(m_k, m_u) = \sum_{n=1}^N c(n) [m_k(n) - m_u(n)]^2, \tag{16}$$

其中, $k, u \leq N_h$. 令 L 是用户选定的隐含层激活函数输出 $O_p(k)$ 的均值向量 $\mathbf{m}_k$ 的数量, $L \leq N_h - 1$, 则忽略零值,按照从小到大的顺序重排 $\mathbf{D}_m$ 的每一行元素,并将 $\mathbf{D}_m$ 的每一行元素中第 L 个最小的元素形成一个新的向量 $\mathbf{d}$. 令 DM 向量 $\mathbf{d}$ 的各个元素 $d(k)$ 满足如下关系:

$$e^{-\beta(k)d(k)} = e^{-1}, \tag{17}$$

则覆盖参数向量 β 的第 k 个元素可以初始化为

$$\beta(k) = \frac{1}{d(k)}, \tag{18}$$

其中, $1 \leq k \leq N_h$. 这种初始化方法的好处是可以保证覆盖系数 $\beta(k)$ 的初始化值既不太大也不太小,从几何意义上说,也保证了各个高斯函数既不太陡也不太平坦. 在本文中,我们根据经验选择 $L = \frac{N_h}{4}$.

2.3 距离测量加权向量 $\mathbf{c}$ 的优化

为了更新距离测量加权向量 $\mathbf{c}$,本文引入了偏差向量 $\mathbf{d}_e$,这样隐含层单元的输入函数如下所示:

$$net_p(k) = \sum_{n=1}^N (c(n) + d_e(n)) [x_p(n) - m_k(n)]^2. \tag{19}$$

在输出向量 $\mathbf{y}_p$ 和 E 沿用上述函数的情况下, $\mathbf{d}_c$ 的一阶梯度可用式(20)计算得到:

$$g_c(n) = -\frac{\partial E}{\partial d_c(n)} = -\frac{2}{N_{vp}} \sum_{p=1}^{N_{vp}} \sum_{i=1}^M [t_p(i) - y_p(i)] \frac{\partial \hat{y}_p(i)}{\partial d_c(n)}, \quad (20)$$

其中:

$$\frac{\partial y_p(i)}{\partial d_c(n)} = -\sum_{k=1}^{N_h} \omega_o(i, k) O_p(k) \times (-\beta(k)) [x_p(n) - m_k(n)]^2. \quad (21)$$

$\mathbf{d}_c$ 的 Hessian 矩阵元素可以用式(22)算出:

$$h_c(u, v) = \frac{\partial^2 E}{\partial d_c(u) \partial d_c(v)} = \frac{2}{N_{vp}} \sum_{p=1}^{N_{vp}} \left[\sum_{i=1}^M \frac{\partial y_p(i)}{\partial d_c(u)} \times \frac{\partial y_p(i)}{\partial d_c(v)} \right]. \quad (22)$$

Newton 法求解 $\mathbf{d}_c$ 的线性方程组表示如下:

$$\mathbf{H}_c \cdot \mathbf{d}_c = \mathbf{g}_c. \quad (23)$$

由于求解式(23)线性方程组中 Hessian 矩阵可能存在奇异性, 直接采用 Gauss-Jordan 消去法会很不安全. 因此, 梯度下降法、共轭梯度算法(conjugate gradient, CG)和正交最小二乘法(orthogonal least square, OLS)等成为更理想的解法. 其中, OLS 法由于具有单步最优化以及在不降低收敛性能的情况下可以减低二阶 Newton 法计算复杂性的能力, 成为最佳的选择. OLS 等价于 QR 分解法. 在本文中, 所有 Newton 法的线性方程组均采用 OLS 法. 这样 OLS 法求解 $\mathbf{d}_c$ 后, $c(n)$ 就可更新为

$$c(n) \leftarrow c(n) + d_c(n). \quad (24)$$

2.4 覆盖参数向量优化

为了实现对第 k 个隐含单元的覆盖参数向量 $\boldsymbol{\beta}$ 的优化, 本文同样引入一个 N_h 维覆盖参数偏置向量 $\mathbf{d}_\beta$. 这样, 对第 k 个隐含单元的覆盖参数更新为

$$\beta_k(k) \leftarrow \beta_k(k) + d_\beta(k), \quad (25)$$

而第 k 个隐含单元输出函数改写为

$$O_p(k) = e^{-(\beta(k) + d_\beta(k)) \cdot \text{net}_p(k)}. \quad (26)$$

这种情况下, $\mathbf{d}_\beta$ 各个元素的梯度可计算为

$$g_\beta(k) = -\frac{\partial E}{\partial d_\beta(k)} = -\frac{2}{N_{vp}} \sum_{p=1}^{N_{vp}} \sum_{i=1}^M [t_p(i) - y_p(i)] \frac{\partial y_p(i)}{\partial d_\beta(k)}. \quad (27)$$

而 $\mathbf{d}_\beta$ 的 Hessian 矩阵为

$$h_\beta(u, v) = \frac{\partial^2 E}{\partial d_\beta(u) \partial d_\beta(v)} = \frac{2}{N_{vp}} \sum_{p=1}^{N_{vp}} \left[\sum_{i=1}^M \frac{\partial y_p(i)}{\partial d_\beta(u)} \times \frac{\partial y_p(i)}{\partial d_\beta(v)} \right]. \quad (28)$$

最优 $\mathbf{d}_\beta$ 同样可以通过采用 OLS 法求解 Newton 法的下列线性方程组算得:

$$\mathbf{H}_\beta \cdot \mathbf{d}_\beta = \mathbf{g}_\beta. \quad (29)$$

2.5 对均值向量 $\mathbf{m}_k$ 的优化

为了学习第 p 个输入样本第 k 个隐含层单元均值向量 $\mathbf{m}_k$, 本算法引入均值向量学习因子向量 $\mathbf{z}$ 并采用多优化学习因子(multiple optimal learning factors, MOLFs)法^[30-31]对 $\mathbf{z}$ 进行学习. 在这种情况下, $\text{net}_p(k)$ 改写为

$$\text{net}_p(k) = \sum_{n=1}^N c(n) \times [x_p(n) - (m_k(n) + z_k g_{\text{mean}}(k, n))]^2, \quad (30)$$

$$g_{\text{mean}}(k, n) = -\frac{\partial E}{\partial m_k(n)} =$$

$$\frac{2}{N_{vp}} \sum_{p=1}^{N_{vp}} \sum_{i=1}^M [t_p(i) - y_p(i)] \frac{\partial y_p(i)}{\partial m_k(n)}. \quad (31)$$

式(30)中, z_k 对应第 k 个隐含单元均值的学习因子. 这样, $\mathbf{z}$ 各个元素的梯度可以写为

$$g_z(k) = -\frac{\partial E}{\partial z_k} = \frac{2}{N_{vp}} \sum_{p=1}^{N_{vp}} \sum_{i=1}^M [t_p(i) - y_p(i) - \omega_{oi}(i, N+2) \sum_{n=1}^N b(n) x_p(n)] \frac{\partial y_p(i)}{\partial z_k}, \quad (32)$$

其中

$$\begin{aligned} \frac{\partial y_p(i)}{\partial z_k} = & -2\omega_o(i, k) O'_p(k) \beta(k) \times \\ & \left[\sum_{n=1}^N c(n) [x_p(n) - m_k(n)] g_{\text{mean}}(k, n) + \right. \\ & \left. c(N+1) \left[\sum_{n=1}^{N+1} b(n) x_p(n) - m_k(N+1) \right] \times \right. \\ & \left. g_{\text{mean}}(k, N+1) \right]. \end{aligned} \quad (33)$$

$\mathbf{z}$ 的 Hessian 矩阵 $\mathbf{H}_z$ 的各个元素为

$$h_z(u, v) = \frac{\partial^2 E}{\partial z_u \partial z_v} \approx$$

$$\frac{2}{N_{vp}} \sum_{p=1}^{N_{vp}} \left[\sum_{i=1}^M \frac{\partial y_p(i)}{\partial z_u} \times \frac{\partial y_p(i)}{\partial z_v} \right]. \quad (34)$$

这样给定 $\mathbf{z}$ 负梯度向量 $\mathbf{g}_z = \left[-\frac{\partial E}{\partial z_1}, -\frac{\partial E}{\partial z_2}, \right.$

$\left. \dots, -\frac{\partial E}{\partial z_{N_h}} \right]^T$ 和 Hessian 矩阵, 优化 $\mathbf{z}$ 即可采用下面的线性方程组求得:

$$\mathbf{z} = \mathbf{H}_z^{-1} \cdot \mathbf{g}_z. \quad (35)$$

从而, 均值向量在每次迭代后可以更新如下:

$$m_k(n) \leftarrow m_k(n) + z_k \cdot g_{\text{mean}}(k, n). \quad (36)$$

2.6 输出权值的优化

基于第 k 个隐含单元和第 m 个输出单元的输出权值 $\mathbf{W}_o$ 的梯度为

$$\hat{g}_o(m,k) = \frac{\partial E}{\partial w_o(m,k)} =$$
$$-2[c_c(m,k) - \sum_{i=1}^{N_{un}} w_o(m,i)r(i,k)]. \quad (37)$$

其中, $r(k,i)$ 为自相关矩阵 $\mathbf{R}$ 的元素并具有以下形式:

$$r(k,i) = \frac{1}{N_{vp}} \sum_{p=1}^{N_{vp}} \bar{x}_p(k) \bar{x}_p(i).$$

(38)

交叉相关矩阵(crossing correlation matrix) $\mathbf{C}_c$ 的元素有以下形式:

$$c_c(m,k) = c_c(k,m) = \frac{1}{N_{vp}} \sum_{p=1}^{N_{vp}} \bar{x}_p(k) t_p(m).$$

(39)

令 $\hat{g}_o(m,k)=0$, 则第 m 个方程组的第 k 个方程为

$$c_c(k,m) = \sum_{i=1}^{N_{un}} r(k,i) w_o(m,i).$$

(40)

综合式(40)和式(37), $\mathbf{W}_o$ 可以通过 OLS 求解下列线性方程组得到:

$$\mathbf{C}_c = \mathbf{R} \cdot \mathbf{W}_o.$$

(41)

3 算法描述

算法 1. 混合优化加权 DM-RBF 的训练算法.

输入: 训练数据集 $\{\mathbf{x}_p, \mathbf{t}_p | 1 \leq p \leq N_{vp}\}$ 和自选最小数 L ;

输出: 多参数优化 RBF 神经网络.

Step1. 读取数据文件获得输入向量, 并规格化输入向量为零均值和单位方差;

Step2. 利用 SOM 法初始化均值向量参数 $\mathbf{m}_k$, 并利用式(15)初始化加权 DM 系数向量 $\mathbf{c}$ 以及式(18)初始化覆盖参数向量 $\boldsymbol{\beta}$;

Step3. 在训练算法每次迭代中, 利用 Newton 法分别优化更新加权 DM 系数向量 $\mathbf{c}$ (见式(20)~(24))、覆盖参数向量 $\boldsymbol{\beta}$ (见式(25)~(29)) 以及均值向量参数 $\mathbf{m}_k$ (见式(31)~(36));

Step4. 用 OWO 求解最优输出权值向量 (见式(39)~(41));

Step5. 返回 Step3, 直至达到指定迭代次数或 MSE 足够小时, 算法结束.

4 实验结果

本文采用“twod”数据集^[32]来分析混合加权 DM-RBF 训练算法的收敛性能. “twod”数据集是一

个高相关性的数据集, 包含 1 768 个样本. 每个样本有 8 个输入单元、7 个输出单元.

为了规范表示各种参数优化算法, 本文用 $\{N(\text{param1}[, \text{param2}][, \text{param3}]) + \text{param4} + [\text{param5}][+ \text{param6}]\}$ 表示不同类型的算法, 其中 param1 , param2 和 param3 标记采用 Newton 法进行训练的参数, 而 param4 , param5 和 param6 标记那些固定采用初始化阶段值并不再参与 RBF 网络训练过程的参数. 例如 $\{N(\mathbf{c}) + \boldsymbol{\beta} + \mathbf{m}\}$, $N(\mathbf{c})$ 表示只对加权 DM 系数向量 $\mathbf{c}$ 用 Newton 法进行了优化, 而 $\boldsymbol{\beta}$ 和 $\mathbf{m}$ 分别表示覆盖参数向量和均值矩阵在初始化阶段固定下来而在训练过程中保持不变.

本文中初始化簇、初始化参数和输入权值均采用同一种方式, 这样各类算法的 MSE 在第 1 次迭代后都是相同, 从而保证不同算法收敛性比较的公平性. 此外, 本文中训练迭代次数均设为 50 次.

4.1 加权 DM 系数的影响

从图 2 可以看出引进加权距离测量 DM 对不同参数训练的 RBF 训练的误差收敛性具有如下影响结果:

- 1) 从曲线 $\{N(\boldsymbol{\beta}) + \mathbf{m}\}$, $\{N(\mathbf{m}) + \boldsymbol{\beta}\}$ 和 $\{N(\boldsymbol{\beta}, \mathbf{m})\}$ 可以看出, 在没有引进加权 DM 系数的情况下, 单纯训练均值参数向量而保持覆盖参数向量不变, 训练效果最佳. 这是因为一旦宽度固定下来, 优化均值向量参数就可以保证训练过程只在高激活的区域进行. 优化 $\mathbf{m}_k$ 本质上是把固定宽度的高斯函数的“帽”覆盖整个权值空间, 从而提高了泛化性.
- 2) 分别比较曲线 $\{N(\boldsymbol{\beta}) + \mathbf{m}\}$ 和 $\{N(\boldsymbol{\beta}) + \mathbf{m} + \mathbf{c}\}$, $\{N(\mathbf{m}) + \boldsymbol{\beta}\}$ 和 $\{N(\mathbf{m}) + \boldsymbol{\beta} + \mathbf{c}\}$ 以及 $\{N(\boldsymbol{\beta}, \mathbf{m})\}$ 和

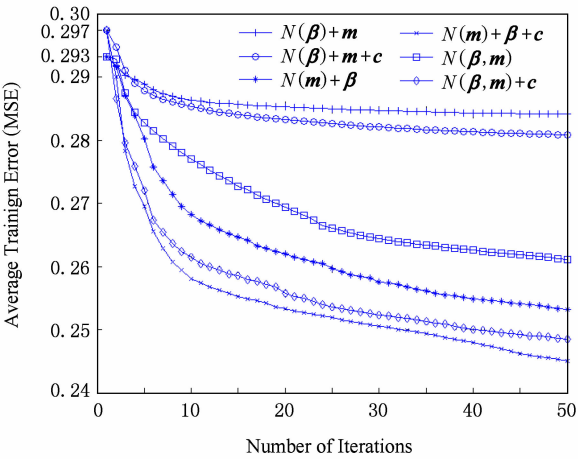


Fig. 2 Effect of weighted Euclidean DM on training different parameters of RBF neural network.
图 2 加权距离测量 DM 对 RBF 网络训练的影响

$\{N(\beta, m) + c\}$, 可以看出尽管引入加权 DM 系数后起始训练误差很大, 但收敛速度比未使用加权 DM 的算法要快. 这也证实了本文对加权 DM 优势的理论解释.

因此, 使用加权 DM 与传统距离测量相比是一个更好的选择.

3) 从所有线图的梯度变化中可以看出, 单纯训练 β 的算法不会有一个明显的收敛性. 这意味着单纯改变 β 的算法不是提高算法性能的好方法.

4.2 Newton 法训练不同 RBF 网络

Newton 法训练不同参数对加权距离测量 DM 算法性能的影响如图 3 所示:

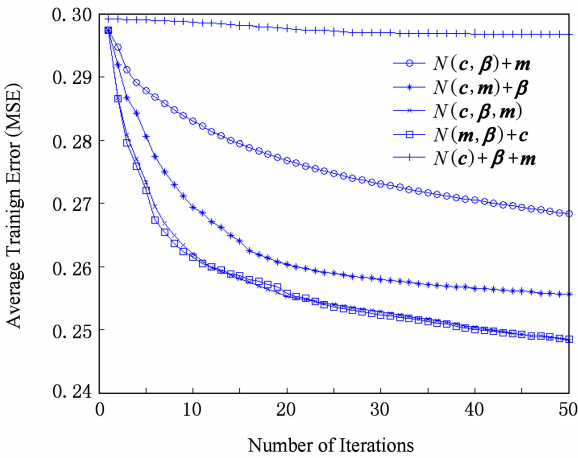


Fig. 3 Effect of using Newton's method on training different parameters for "twod" dataset.

图 3 在“twod”数据集上对不同的参数应用 Newton 法的影响

图 3 显示了牛顿法在不同参数优化上应用的结果, 可以得到下面观察结果:

1) 如曲线 $\{N(c) + \beta + m\}$ 所示, 单纯应用 Newton 法在 DM 加权系数上并不能很好提高收敛性能. 此外, 通过比较 $\{N(c, \beta) + m\}$ 和 $\{N(c, m) + \beta\}$ 可以看出, 加权系数的训练对改变性能作用不大. 这是因为采用高斯函数的宽度(由传播参数控制)和顶点位置(均值向量)并没有完全被优化, 因此导致不理想的收敛性能.

2) 均值或传播参数其中之一或二者都应用 Newton 法时(除 $\{N(c) + \beta + m\}$ 以外的其他 4 条曲线所示), 可以看出收敛率大大加快.

3) 比较曲线 $\{N(c, \beta) + m\}$ 以及 $\{N(c, m) + \beta\}$, 可以看出 Newton 法优化均值向量对收敛性提高影响最大.

综上所述, 加权系数向量优化只有和均值向量、覆盖(传播)参数优化结合, 才可以得到更好的 RBF 网络.

4.3 Newton 法训练加权 DM-RBF 混合网络

在传统的训练算法中加入加权 DM 后, 我们一是希望看到加权 DM 与传统网络参数合并后的效果, 二是希望看到加权 DM 混合参数应用 Newton 法优化后的效果. 图 4 表示出了关于应用 Newton 法进行混合参数优化后的 4 个重要的事实.

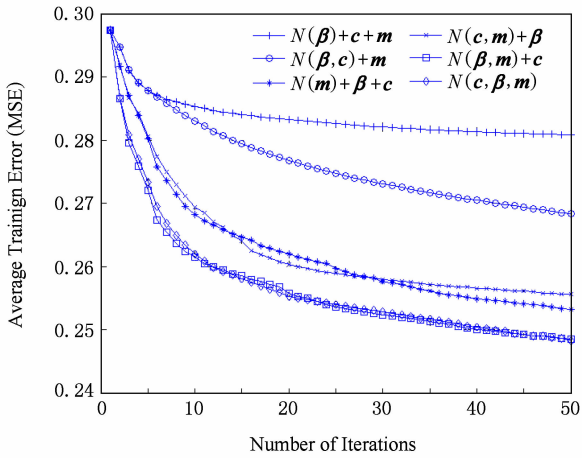


Fig. 4 Effect of Newton's method on various hybrid weighted DM-RBF training algorithms.

图 4 Newton 法对各混合加权 DM-RBF 训练算法影响

从图 4 得出以下结论:

1) 比较 $\{N(\beta) + c + m\}$ 和 $\{N(\beta, c) + m\}$ 可以看出, 单纯应用 Newton 法于覆盖参数, 误差曲线没有显著差异, 然而混合了优化加权 DM 系数后, 覆盖参数的收敛性大大提高.

2) 比较 $\{N(m) + \beta + c\}$ 和 $\{N(c, m) + \beta\}$ 可以看出, 均值向量参数 m_k 对网络的整体表现起主导作用, 而且如果在此基础上对加权 DM 系数上应用 Newton 法, 则可以产生光滑的收敛行为.

3) 从曲线 $\{N(c, \beta, m)\}$ 和 $\{N(\beta, m) + c\}$ 可以看出, 在混合了优化覆盖参数和优化均值向量后, RBF 网络没有显著提升.

比较所有曲线, 优化加权 DM 系数可以大大提高优化均值向量及覆盖参数的性能, 因此是一种比较理想的 RBF 训练算法.

4.4 混合加权 DB-RBF 与传统 RBF 算法性能比较

为了分析混合加权 DM-RBF 的性能, 本文在“twod”数据集上比较了相同初始化数据条件下加权 DM-RBF, LM-RBF 以及 RLS-RBF (recursive

least square RBF)3 种训练算法的误差收敛率,结果如图 5 所示:

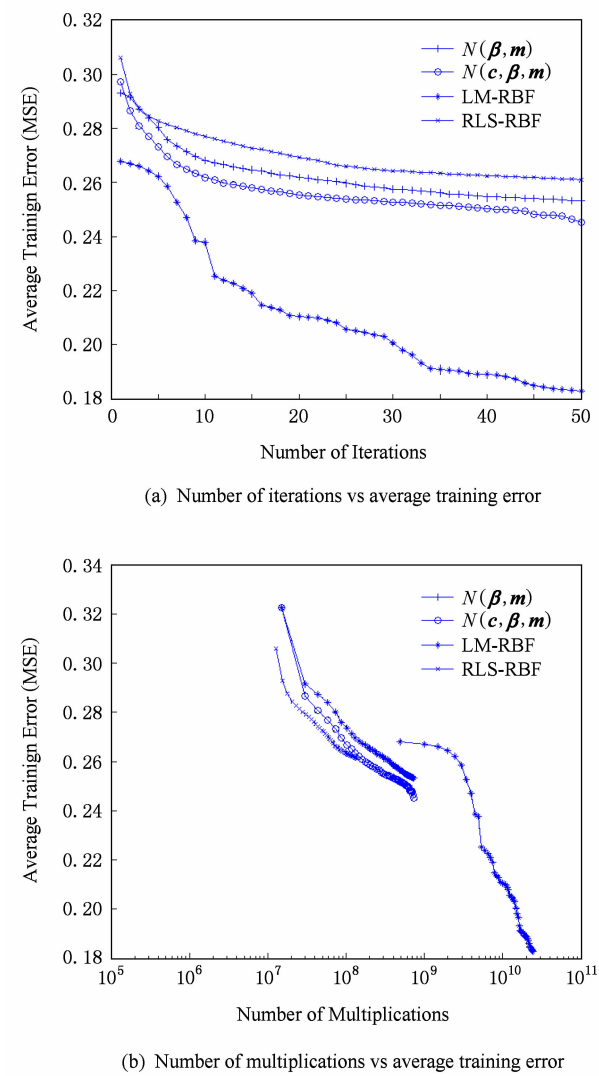


Fig. 5 Performance of weighted DM-RBF training algorithm compared with LM-RBF algorithm and RLS-RBF algorithm.
图 5 加权 DM-RBF 与 LM-RBF, RLS-RBF 训练算法的性能比较

从图 5 可以看出:

1) 通过比较曲线 $\{N(\beta, m)\}$ 和 $\{N(c, \beta, m)\}$ 可以看出, 加权 DM 改进了非加权 DM 的算法性能;

2) 从图 5(a) 可以看出基于 LM 算法的 LM-RBF 网络训练算法比其他 2 种训练算法的收敛性都好; 然而从图 5(b) 看出 LM-RBF 算法的乘运算次数比其他 2 种训练算法多得多. 因此计算复杂度限制了其在大数据集上的实际应用.

综上所述, 本文所提出的算法具有在计算复杂度较低的同时保持良好的训练性能的特点, 因而更适合于大规模数据集.

5 结束语

本文主要提出一种加权 DM 的混合 RBF 训练算法, 并通过大量实验验证了如下结论:

1) 加权 DM 大大提升 RBF 网络收敛性能;

2) 采用 Newton 法调整均值向量参数以及调整传播参数比调整加权系数向量更重要;

3) 混合优化加权 DM 系数可以大大提高单纯优化网络参数向量的算法性能;

4) Newton 法优化的加权 DM-RBF 比传统 RBF 训练算法有效而且易于实现, 因而具有更好的应用价值.

未来的工作将研究如何可以根据不同的数据集优化选择隐含层单元数量的方法, 从而进一步提高网络训练收敛性和泛化性.

参 考 文 献

[1] Duda R O, Hart P E. Pattern Classification and Scene Analysis [M]. New York: Wiley, 1973

[2] Poggio T, Girosi F. A Theory of Networks for Approximation and Learning [M]. Massachusetts: Artificial Intelligence Laboratory and Center for Biological Information Processing and Whitaker College, Massachusetts Institute of Technology, 1989: 11-45

[3] Powell M J D. Radial basis functions for multivariate interpolation: A review [M] // Algorithms for the Approximation of Functions and Data. New York: Clarendon Press, 1987: 143-167

[4] Bhowmik P S, Pradhan S, Prakash M, et al. Investigation of wavelets and radial basis function neural network for incipient fault diagnosis in induction motors [C] // Proc of the 3rd IEEE CCUBES. Los Alamitos, CA: IEEE Computer Society, 2013: 27-28

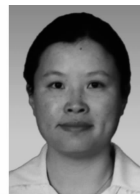
[5] Alexandridis A, Chondrodima E, Sarimveis H. Radial basis function network training using a nonsymmetric partition of the input space and particle swarm optimization [J]. IEEE Trans on Neural Networks and Learning Systems, 2013, 24 (2): 219-230

[6] Manning T, Walsh P. Improving the performance of cgpnn for breast cancer diagnosis using crossover and radial basis functions [G] // LNCS 7833: Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics. Berlin: Springer, 2013: 165-176

[7] Mohsen H, Abbas R, Leila N. Application of radial basis function network for the modeling and simulation of turbogenerator [J]. Journal of Advances in Information Technology, 2013, 4(2): 76-79

[8] Yin Jianchuan, Zou Zaojian, Xu Feng. Sequential learning radial basis function network for real-time tidal level predictions [J]. Ocean Engineering, 2013, 57(1): 49-55

- [9] Mehraia A, Karimib H R, Thobena K D, et al. Application of learning pallets for real-time scheduling by the use of radial basis function network [J]. *Neurocomputing*, 2013, 101 (4): 82-93
- [10] Ge Yanqiang, Wang Xiangzheng, Yu Jiangde. A method of neural blocking forecasting about Kohonen self-organizing Maps and radial basis function network [J]. *Microelectronic & Computer*, 2012, 29(12): 176-179 (in Chinese)
(葛彦强, 汪向征, 于江德. 一种利用自组织映射和径向基函数神经网络的网络拥塞预测方法[J]. *微电子学与计算机*, 2012, 29(12): 176-179)
- [11] Zhu Fuzhen, Li Jinzong, Zhu Bing, et al. Super-resolution image reconstruction based on RBF neural network [J]. *Optics and Precision Engineering*, 2010, 18(6): 1444-1451 (in Chinese)
(朱福珍, 李金宗, 朱兵, 等. 基于径向基函数神经网络的超分辨率图像重建[J]. *光学精密工程*, 2010, 18(6): 1444-1451)
- [12] Moddy J E, Darken C J. Fast learning in networks of locally-tuned processing units [J]. *Neural Computation*, 1989, 1 (2): 281-294
- [13] Cha I, Kassam S A. Interference cancellation using radial basis function network. *Signal Processing*, 1995, 47 (3): 247-268
- [14] Battiti R. First and second order methods of learning: between steepest descent and Newton method [J]. *Neural Computation*, 1992, 4(2): 141-166
- [15] Wasserman P D. *Advanced Methods in Neural Computing* [M]. New York: John Wiley & Sons Inc, 1993
- [16] Karayiannis N B. Gradient descent learning of radial basis neural networks [C] //Proc of the 7th IEEE Int Conf on Neural Networks. Houston: Institute of Electrical and Electronics Engineers Inc, 1997: 1815-1820
- [17] Marquardt D. An algorithm for least-squares estimation of nonlinear parameters [J]. *Society for Industrial and Applied Mathematics*, 1963, 11(2): 431-441
- [18] Shi Yixun. Globally convergent algorithms for unconstrained optimization [J]. *Computational Optimization and Application*, 2000, 16(3): 295-308
- [19] Zhang Long, Li Kang, Bai Erwei. A new extension of newton algorithm for nonlinear system modelling using rbf neural networks [J]. *IEEE Trans on Automatic Control*, 2013, 58(11): 2929-2933
- [20] Huang Guangbin, Saratchandran P, Sundararajan N. A generalized growing and pruning RBF (GGAP-RBF) neural network for function approximation [J]. *IEEE Trans on Neural Networks*, 2005, 16(1): 57-67
- [21] Chen Sheng, Cowan C F N, Grant P. M. Orthogonal least squares learning algorithms for radial basis functions networks [J]. *IEEE Trans on Neural Network*, 1991, 2(2): 302-309
- [22] Chen Sheng, Chng E S, Alkadhimi K. Regularized orthogonal least square algorithm for construction radial basis function networks [J]. *International Journal of Control*, 1996, 64(5): 829-837
- [23] Chng E S, Chen Sheng, Mulgrw B. Gradient radial basis function network for nonlinear and nonstationary time series prediction [J]. *IEEE Trans on Neural Networks*, 1996, 7 (1): 190-194
- [24] Orr M J L. Regularization in the selection of radial basis function centers [J]. *Neural Computation*, 1995, 7(3): 606-623
- [25] Wang Xuelei, Shao Huihe, Li Yafen. An online training algorithm for RBF neural network and its application in nonlinear control [J]. *Information and Control*, 2001, 30 (3): 249-252 (in Chinese)
(王学雷, 邵惠鹤, 李亚芬. 一种径向基函数神经网络在线训练算法及其在非线性控制中的应用[J]. *信息与控制*, 2001, 30(3): 249-252)
- [26] Tan Jianhui. Relearning algorithm and application of radial basis function neural network [J]. *Microelectronics & Computer*, 2006, 23(5): 115-120 (in Chinese)
(谭建辉. 径向基函数神经网络的再学习算法及其应用[J]. *微电子学与计算机*, 2006, 23(5): 115-120)
- [27] Barnard E. Optimizations for training neural nets [J]. *IEEE Trans on Neural Networks*, 1992, 3(2): 232-240
- [28] Haykin S. *Neural Network: A Comprehensive Foundation* [M]. 3rd ed. Englewood Cliffs, NJ: Prentice Hall, 2009
- [29] De Wachter M, Demuyneck K, Wambacq P, et al. A locally weighted distance measure for example based speech recognition [C] //Proc of the 29th IEEE Int Conf on Acoustic, Speech and Signal Processing. Los Alamitos, CA: IEEE Computer Society, 2004: 1-181-184
- [30] Malalur S S, Manry M T, Jesudhas P. Multiple optimal learning factors for the multi-layer perceptron [J]. *Neurocomputing*, 2015, 149: 1490-1501
- [31] Cai, Xun, Manry M, Tyagi K. Training multilayer perceptron by using optimal input normalization [C] //Proc of the 24th IEEE Int Conf on Fuzzy Systems. Los Alamitos, CA: IEEE Computer Society, 2011: 2771-2778
- [32] Image Processing and Neural Networks Lab. Training data files for regression or approximation [EB/OL]. (2009-06-02) [2010-06-02]. http://www.uta.edu/faculty/manry/new_mapping.html



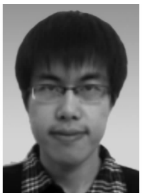
Cai Xun, born in 1971. PhD and associate professor in Shandong University. Senior member of China Computer Federation. Her main research interests include artificial neural network, image recognition and digital signal processing.



Chen Zhi, born in 1990. Master from Shandong University. His main research interests include machine learning and deep models(chenzhisdu@gmail.com).



Kanishka Tyagi, born in 1984. PhD candidate from the University of Texas at Arlington. His main research interests include artificial neural network, machine learning and digital signal processing (kanishka.tyagi@mavs.uta.edu).



Li Ziqiang, born in 1991. Master candidate in Shandong University. Student member of China Computer Federation. His main research interest is artificial neural network(alex_li@acm.org).



Yu Kuan, born in 1975. Received his PhD degree from Shandong University. Currently ssociate professor in Shandong Jianzhu University. His main research interests include computer control and simulation of high performance material PAN and carbon fiber production(yukuan@sdjzu.edu.cn).



Zhu Bo, born in 1969. Received his PhD degree from the University of Science and Technology Beijing. Currently professor and PhD supervisor in Shandong University. His main research interests include computer control and simulation of high performance material PAN and carbon fiber production(zhubo@sdu.edu.cn).

2016 年《计算机研究与发展》专题(正刊)征文通知
——数据融合与知识融合

随着大数据时代的到来,如何从海量异构的数据源中抽取知识,并进行知识的融合,从而构建大规模的知识图谱,是目前数据管理和信息抽取领域的一个重要学术课题.相比于传统的数据集成,在面向大规模的数据和知识融合过程中的,融合算法的效率、多源数据的数据质量评估和基于语义的数据与知识融合等都给现有的数据集成技术带来了巨大的挑战.

《计算机研究与发展》将于 2016 年 2 月出版数据融合与知识融合专题,侧重大规模数据和知识的抽取、融合和应用等方面,将涉及到数据管理、信息抽取和知识库构建等多个交叉学术领域,这为相关领域的研究者提供更多的沟通和交流的机会,共同探讨面向大数据的数据融合与知识融合的基础理论研究及其应用,讨论该领域内最新的突破性进展和未来的发展趋势.

欢迎相关领域的专家学者和科研人员踊跃投稿.现将专题论文征集的有关事项通知如下.

征文范围(包括但不限于如下内容)

- 数据和知识抽取技术

• 实体和关系抽取

• 事件抽取

—异构性消除

• 实体消歧和实体链接

• 本体构建

• 本体匹配

• 知识库扩充

—数据与知识融合技术
- 数据清理

• 数据冲突

• 数据源质量评估

• 数据源关联挖掘

(copy/correlation detection)

• 机器学习与知识融合

• 构建的知识库质量评估

—数据和知识建模

• 数据和知识概率化
- 文本知识化建模

• 语义计算

• 多语言的知识库

—知识融合的应用

• 知识库检索

• 基于知识库的信息抽取和概念理解

• 基于知识的文本理解

• 基于知识的智能应用

• 基于知识的隐私保护

征文要求

- 1) 论文应属于作者的科研成果,数据真实可靠,具有重要的学术价值与推广应用价值,未在国内外公开发行的刊物或会议上发表,不存在一稿多投问题.作者在投稿时,需向编辑部提交版权转让协议.
- 2) 论文一律用 word 格式排版,论文格式体例参考近期出版的《计算机研究与发展》的要求(<http://crad.ict.ac.cn/>).
- 3) 论文通过期刊网站(<http://crad.ict.ac.cn/>)投稿,投稿时提供作者的联系方式,作者留言中务必注明“数据融合 2016 专题”(否则按自由来稿处理).

重要日期

征文截止日期: 2015 年 9 月 15 日

录用通知日期: 2015 年 10 月 15 日

作者修改稿提交日期: 2015 年 11 月 5 日

出版日期: 2016 年 2 月

特邀编委

孟小峰 教授 中国人民大学 xfmeng@ruc.edu.cn

王建勇 教授 清华大学 jianyong@tsinghua.edu.cn

董欣 高级研究员 Google Research

联系方式

编辑部 crad@ict.ac.cn 010-62620696, 010-62600350

通信地址: 北京 2704 信箱《计算机研究与发展》编辑部 邮政编码: 100190