

用于风电功率预测的 RPCL 优化神经网络模型

苏义鑫, 夏慧雯

(武汉理工大学自动化学院, 武汉 430070)

摘要: 为了提高风电功率预测的准确度, 提出了一种基于对手竞争惩罚学习算法 (rival penalized competitive learning, RPCL) 优化径向基函数 (radial basis function, RBF) 神经网络的风电功率预测模型. 首先通过 RPCL 确定网络隐含层神经元数目以及中心点初始值, 然后由 K 均值聚类法确定隐含层神经元的中心点和宽度, 最后通过最小均值算法确定隐含层神经元与输出层神经元之间的权重. 仿真结果表明: 此优化模型相较于传统 RBF 网络具有更高的准确性.

关键词: 风电功率预测; 对手竞争惩罚学习算法; RBF 神经网络; K 均值聚类

中图分类号: TM 614

文献标志码: A

文章编号: 0254-0037(2016)05-0674-05

doi: 10.11936/bjtxb2015110029

Rival Penalized Competitive Learning-based Neural Network Model for Wind Power Forecasting

SU Yixin, XIA Huiwen

(School of Automation, Wuhan University of Technology, Wuhan 430070, China)

Abstract: For increasing the accuracy of wind power forecasting, a rival penalized competitive learning-based radial basis function (RBF) neural network model was presented. Firstly the number of neural network hidden-layer-nodes and its initial center values were determined by rival penalized competitive learning. And then the width of RBF and the center values of network were identified accurately through K -means clustering. At last, appropriate weights of network were estimated by least mean square. The forecasting result shows that the presented model can lead to more accurate forecasting compared with the traditional neural network.

Key words: wind power forecasting; rival penalized competitive learning algorithm; radial basis function neural network; K -means clustering

近年来, 随着化石能源的逐渐枯竭, 风能作为清洁的可再生能源, 得到了大规模的发展. 与此同时国内外关于风电的课题研究也越来越深入, 但其中风电功率预测的精度还没达到令人满意的程度^[1].

由于风力发电主要使用的是近地风能, 风电功率具有很强的随机性, 因此, 大规模的风电场并网会

影响电网的稳定性和供电系统的可靠性. 只有通过准确地预测风电场风电功率, 为电力调度部门及时调整调度计划提供可靠的依据, 才能使电力供需达到平衡, 减小风电场对电网的影响. 同时也会大幅降低电网的备用容量, 减少电力系统运行成本.

目前风电功率的预测方法主要分为物理法、统

收稿日期: 2015-11-10

基金项目: 湖北省自然科学基金资助项目(2015CFB586)

作者简介: 苏义鑫(1965—), 男, 教授, 主要从事智能控制理论与应用、智能系统关键技术方面的研究, E-mail: suyixin@whut.edu.cn

计法(时间序列法、回归分析法等)和学习法(支持向量机、小波分析和神经网络等)。径向基函数(radial basis function, RBF)神经网络具有逼近任意非线性映射的能力,已广泛应用于各种领域,但是其网络结构和参数的选取仍没有系统的方法来确定。本文利用基于对手竞争惩罚学习(rival penalized competitive learning, RPCL)算法的 K 均值聚类对数据进行处理,确定隐含层的中心。其中,RPCL^[2-4]是由 Xu 等在 1993 年提出的次胜单元受到惩罚的学习机制,具有优良的算法性能,能确定数据样本的合理聚类数,现已被广泛应用于各种领域,如图像处理^[5-6]、语音识别^[7]等。然后,通过最小均值算法确定网络权值。最后,利用中国某风电场的风电功率数据进行建模和预测实验,仿真结果显示:该方法具有较高地预测精度。

1 基于 K 均值聚类优化的 RBF 神经网络

RBF 神经网络属于前向网络,由输入层、中间层和输出层组成,其网络拓扑图如图 1 所示。从图 1 可以看到,每层的神经元接受上层传出的信息后单向向下层传输,同层的神经元之间没有信息传输,不同层之间也没有反馈。Kolmogorov 理论表明:具有合适的网络结构和权值的三层前向神经网络可以逼近任意精度的函数^[8]。

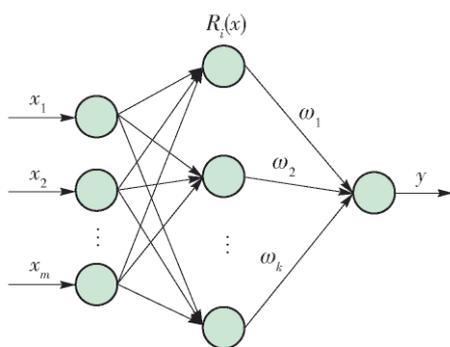


图 1 RBF 神经网络拓扑图
Fig. 1 Topology of RBF neural network

RBF 神经网络的径向基函数一般采用高斯激活函数,其表现形式(也就是隐含层节点的输出)为

$$R_i(x) = \exp\left(-\frac{1}{2\sigma_i^2} \|x - c_i\|^2\right) \quad i = 1, 2, \dots, k \quad (1)$$

式中: m, k, n 分别为网络输入层、隐含层和输出层的节点数(神经元数); $x = \{x_1, x_2, \dots, x_m\}$ 为 m 维的输入样本; $R_i(x)$ 为隐含层第 i 个节点的输出; c_i 为隐含层第 i 个节点高斯函数的中心; σ_i 为隐含层第 i

个节点高斯函数的宽度; $\|x - c_i\|$ 为欧式范数,表示 x 和 c_i 之间的距离。从式(1)可以看出, R_i 的输出范围为 $0 \sim 1$, 当输入样本离节点中心越近,相应节点的输出值就越大,在 $x = c_i$ 时,输出值为 1。

根据 RBF 神经网络的结构,可以得到网络的输出为

$$y = \sum_{i=1}^k w_i \exp\left(-\frac{1}{2\sigma_i^2} \|x - c_i\|^2\right) \quad (2)$$

式中: y 是输出层节点的输出,在本文风电功率预测中只有一个输出; w_i 为隐含层第 i 个节点与输出层节点之间的权值。

RBF 神经网络需要确定 3 个重要参数:基函数的中心 c_i 、第 i 个神经元宽度 σ_i 和输出层的权值 w_i 。现有很多聚类分析方法用来确定网络的中心初始值,其中 K 均值聚类算法使用最为广泛^[9-10]。给定一个数据点集合和聚类数目 K 值(用户自己指定), K 均值聚类算法根据某个距离函数将所有数据点分入 K 个聚类中,能够比较准确地确定聚类中心。但是传统的 K 均值聚类算法存在 2 个重要的问题:一是需要提前给定 K 的数值;二是初始聚类中心对聚类结果影响较大,有可能收敛到局部最优解,而不是全局最优解。本文采用 RPCL 算法来解决 K 均值聚类算法中的 2 个问题:确定合适的聚类数目 K 和理想的初始聚类中心^[11-12]。

2 标准 RPCL 算法

RPCL 算法的基本思想为:每个输入样本将竞争获胜单元引向它的同时将次胜单元推开,最终,算法可自动地确定最佳聚类数目。同时也有不少学者对 RPCL 算法进行了各种改进,如 RPCL 算法在节点距离调节时,输入数据的几何结构会对距离调整产生影响,谢娟英等^[13]将样本分布密度引入进行节点距离调整,结果表明基于密度的 RPCL 算法在大数据集上表现优良,而在小规模数据集上传统的 RPCL 更优。由于风电功率预测所用数据属于小规模数据集,因此,本文使用标准 RPCL 算法来进行聚类。

给定数据集 $X = \{x_1, x_2, \dots, x_N\}$, 其中样本 $x_l = [x_{l1}, x_{l2}, \dots, x_{lm}]$, $l = 1, 2, \dots, N$ 。预先设置有 k 个节点,一个节点表示一个类簇,将类簇的中心记为 c_i , $i = 1, 2, \dots, k$; 节点 i 的输出值为 $u_i \in \{-1, 0, 1\}$ 。当节点 i 为获胜节点时, $u_i = 1$, 被输入样本吸引;当节点 i 为次胜节点时, $u_i = -1$, 被输入样本所排斥;除此之外,其他节点的输出为 $u_i = 0$, 将不受输入样本

的影响. 具体步骤如下:

步骤1 随机初始化 k 个聚类中心 $\{c_i\}_{i=1}^k$, 并且设置迭代步数 $t=1$ 和最大迭代步数 T (聚类中心修改的最大次数).

步骤2 从数据集中随机选取样本 x_l , 有

$$u_i = \begin{cases} 1, & i=s, \gamma_s \|x_l - c_s\|^2 = \min_{i=1}^k \gamma_i \|x_l - c_i\|^2 \\ -1, & i=r, \gamma_r \|x_l - c_r\|^2 = \min_{i=1, i \neq s}^k \gamma_i \|x_l - c_i\|^2 \\ 0, & \text{其他} \end{cases} \quad (3)$$

式中: s 为获胜单元标号; r 为次胜单元标号; γ_i 为第 i 个聚类中心的调整频率, $\gamma_i = n_i / \sum_{v=1}^k n_v$; n_i 为中心点 i 成为获胜节点的累加次数.

步骤3 更新聚类中心 c_i

$$c_i = c_i + \Delta c_i \quad (4)$$

$$\Delta c_i = \begin{cases} \alpha_s (x_i - c_i), & u_i = 1 \\ -\beta_r (x_i - c_i), & u_i = -1 \\ 0, & \text{其他} \end{cases} \quad (5)$$

式中: α_s 和 β_r 分别是竞争获胜单元和次胜单元的学习率, 大小为 $0 \sim 1$, 通常 α_s 远远大于 β_r . 每次修改聚类中心后, 将 c_i 中的数据与阈值 δ_1 进行比较, 这是因为在风电功率预测模型中, 输入数据会先进行归一化处理, 所以这里的 δ_1 应为 1. 若 c_i 中有数据超出 δ_1 , 则此节点无意义, 将其删除.

步骤4 将 t 值加 1, 若算法收敛或 $t > T$, 算法结束; 否则转 2) 继续迭代.

步骤5 统计每个中心成为获胜节点的累加数量, 如果该数值比阈值 δ_2 (用于删除冗余项) 小则删除该节点.

步骤6 输出满足条件的节点数以及相应中心点.

3 RPCL 优化的神经网络预测模型

经过上述分析, RPCL 优化的神经网络模型 (RPCL-RBF) 可以按以下几个基本步骤进行.

步骤1 样本数据进行归一化处理

$$x'_{li} = (x_{li} - v_{\min}) / (v_{\max} - v_{\min}) \quad (6)$$

$$i=1, 2, \dots, m; l=1, 2, \dots, N$$

式中: x_{li} 、 x'_{li} 分别为数据转换前、后的值; v_{\min} 、 v_{\max} 分别为 i 维样本数据的最大值和最小值.

步骤2 利用 RPCL 算法对归一化的样本数据

进行处理, 得到隐含层节点数和相应的中心点.

步骤3 将上一步的输出作为 K 均值聚类算法的最佳聚类数目 K 和初始聚类中心 $c_i(0)$, $i=1, 2, \dots, k$.

步骤4 计算每个样本到各中心的距离

$$d_i = \|x'_l - c_i(u)\| \quad (7)$$

然后将其分配到距离最近的聚类中心

步骤5 重新计算聚类中心:

$$c_i(u+1) = \frac{1}{n} \sum_{x_l \in X'_i} x_l \quad (8)$$

式中: X'_i 为距离该中心点 i 最近的样本点集; n 为该样本集中样本点的个数. 判断新的聚类中心有无变化, 若无变化, 则算法结束; 否则, 转到步骤 4 继续迭代.

步骤6 将 K 均值聚类算法得到的结果作为 RBF 神经网络的中心, K 均值聚类属于自组织选择法, 在基函数为高斯函数时, 可按

$$d_i = \min_{j=1, j \neq i}^k \|c_j - c_i\| \quad (9)$$

$$\sigma_i = \lambda d_i \quad (10)$$

计算网络的基函数宽度. 式中 λ 为重叠系数.

步骤7 采用递推最小均方误差 (least mean square, LMS) 来训练权值 w_i :

$$\Delta w_i(v) = \eta (y - \hat{y}) R_i(x) \quad (11)$$

$$w_i(v+1) = w_i(v) + \Delta w_i(v) \quad (12)$$

式中: y 为期望输出; \hat{y} 为神经网络模型的输出; η 为学习率 (取较小的正数); v 为权值迭代次数; 权值初始值 $w_i(0)$ 为一较小的随机非零值.

步骤8 用训练好的 RBF 神经网络来进行风电功率预测.

4 应用实例

本文采用中国某一风电场某年 5 月 10 日至 6 月 6 日这段时间内风电机组输出功率的历史数据作为学习样本, 不考虑天气影响. 该数据集每隔 15 min 采集 1 次, 将这些数据作为原始时间序列, 利用前 3 个小时 (12 个点) 的数据来预测第 13 个点的数, 如式 (14) 所示.

$$X_t = (x(t), x(t+1), \dots, x(t+11)) \quad (13)$$

$$t=1, 2, \dots, 2677$$

$$f(X_t) = x(t+12) \quad (14)$$

将时间序列的前 27 天的 2592 个数据作为训练数据, 将第 28 天的 96 个数据作为测试数据. 在 Matlab 环境下建立 RPCL-RBF 模型, 然后用测试数

据来验证训练好的网络,实际输出与 RBF 网络预测输出对比如图 2 所示。

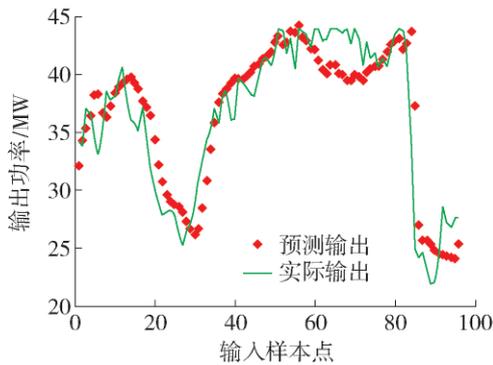


图 2 实际输出与预测输出对比

Fig. 2 Comparison of actual output and predicted output

本文 RPCL-RBF 模型是对 K 均值聚类 RBF 模型神经网络模型的进一步优化. 为了突出经过 RPCL 优化后的模型具有更准确的预测效果,将 RPCL-RBF 模型与 K 均值聚类 RBF 模型的预测结果进行比较,其中为避免其他因素影响,将 K 均值聚类 RBF 选取与 RPCL-RBF 模型相同的隐含层初始个数以及初始中心. 2 个模型的相对误差对比情况如图 3 所示。

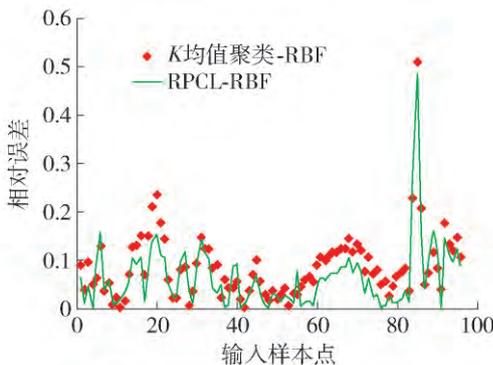


图 3 2 种预测方法的相对误差的比较

Fig. 3 Comparison of absolute error of two forecasting methods

从图 3 可以明显地看出,RPCL-RBF 模型整体的相对误差比 K 均值聚类 RBF 模型的相对误差要小. 为了进一步说明 RPCL-RBF 模型的适用性,本文将其与误差反向传播(back propagation, BP)模型、RBF 模型和 K 均值聚类的 RBF 模型进行对比,为避免其他因素影响,对比的 3 种模型选取与 RPCL-RBF 模型相同的网络结构即隐含层神经元个数相同,并从以下几个方面来对模型进行衡量:首先,从统计的角度由根均方差(root mean square

error, RMSE) 和平均绝对百分误差(mean absolute percentage error, MAPE) 来衡量模型的好坏;然后,从国家能源局发布的风电场功率预测考核指标^[14]即发电预测的准确率 r_1 来衡量预测的可靠性;最后,从算法运行的时间来衡量模型的可用性. 具体数据列于表 1 中. 其中算法运行时间主要包括 2 个部分:离线训练时间和在线预测时间. 但是在模型设计中后者比前者更为重要,这是由于神经网络在离线训练时可以有足够的时间调整参数使其得到最佳模型,因此,本文主要对比的是模型在预测风电功率时所用的时间。

表 1 各项指标的比较

Table 1 Comparison of indicators

模型	RMSE/ MW	MAPE/ %	r_1 / %	t / ms
BP	4.524	9.79	90.82	14.02
RBF	3.568	8.36	92.76	15.54
K 均值聚类 RBF	3.013	6.76	93.89	17.01
RPCL-RBF	2.841	6.36	94.24	12.37

从表 1 可以看出,与常用的几种神经网络模型相比,RPCL-RBF 模型的 RMSE 和 MAPE 值最小,准确率 r_1 最高,且预测所用时间最短。

5 结论

1) 采用了一种由 RPCL 和 K 均值聚类算法组合优化的 RBF 神经网络模型进行风电功率预测. 相比于 K 均值聚类 RBF 模型,RPCL-RBF 模型可以先确定合适的聚类数目 K 和理想的初始聚类中心,使得预测结果更加精确。

2) 通过对中国某风电场的实际运行数据进行建模和预测实验,RPCL-RBF 算法的准确度高于常用的 BP 算法、RBF 算法和 K 均值聚类 RBF 算法,且用时更少,证明了本文方法的可行性和有效性。

参考文献:

[1] 王丽婕,廖晓钟,高阳,等. 风电场发电功率的建模与预测研究综述[J]. 电力系统保护与控制, 2009, 37(13): 118-121.
WANG L J, LIAO X Z, GAO Y, et al. Summarization of modeling and prediction of wind power generation [J]. Power System Protection and Control, 2009, 37(13): 118-121. (in Chinese)

[2] AHALT S C, KRISHNAMURTHY A K, CHEN P, et al.

- Competitive learning algorithms for vector quantization [J]. *Neural Networks*, 1990, 3(3): 277-290.
- [3] XU L, KRZYŻAK A, OJA E. Rival penalized competitive learning for clustering analysis, RBF net, and curve detection [J]. *IEEE Transactions on Neural Networks*, 1993, 4(4): 636-649.
- [4] CHEUNG Y M. On rival penalization controlled competitive learning for clustering with automatic cluster number selection [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2005, 17(11): 1583-1588.
- [5] 李涛, 裴文江, 王少华, 等. 竞争与动态合作学习聚类分析算法[J]. *哈尔滨工程大学学报*, 2010, 31(1): 102-108.
LI T, PEI W J, WANG S H, et al. Competitive and dynamic cooperative learning algorithm [J]. *Journal of Harbin Engineering University*, 2010, 31(1): 102-108. (in Chinese)
- [6] 李桂枝, 安成万, 张永谦, 等. 基于模糊熵和 RPCL 的彩色图像聚类分割[J]. *中国图像图形学报*, 2005, 10(10): 1264-1268.
LI G Z, AN C W, ZHANG Y Q, et al. Color image clustering segmentation based on fuzzy entropy and RPCL [J]. *Journal of Image and Graph*, 2005, 10(10): 1264-1268. (in Chinese)
- [7] 李昕, 费敏锐. 用改进的 RPCL 和 EM 算法确定 EBF 网络结构和参数的策略及其应用[J]. *模式识别与人工智能*, 2003, 16(2): 204-207.
LI X, FEI M R. Combining improved RPCL and EM algorithm to determine the EBF architecture and parameters with its application [J]. *Pattern Recognition and Artificial Intelligence*, 2003, 16(2): 204-207. (in Chinese)
- [8] HORNICK K, STICHCOMBE M, WHITE H. Multilayer feed forward networks are universal approximators [J]. *Neural Network*, 1989, 2: 359-366.
- [9] 张彬. 基于聚类算法的 RBF 神经网络设计综述[J]. *微型机与应用*, 2012, 31(12): 1-3.
ZHANG B. Overview on design of RBF network based on fuzzy clustering [J]. *Microcomputer & Its Applications*, 2012, 31(12): 1-3. (in Chinese)
- [10] XIE J Y, JIANG S, XIE W X, et al. An efficient global K-means clustering algorithm [J]. *Journal of Computers*, 2011, 6(2): 271-279.
- [11] 谢娟英, 郭文娟, 谢维信, 等. 基于样本空间分布密度的初始聚类中心优化 K 均值算法[J]. *计算机应用研究*, 2012, 29(3): 888-892.
XIE J Y, GUO W J, XIE W X, et al. K-Means clustering algorithm based on optimal initial centers related to pattern distribution of samples in space [J]. *Application Research of Computers*, 2012, 29(3): 888-892. (in Chinese)
- [12] 沈佳杰, 江红, 王肃. 聚类中心计数值和全局距离变量 RPCL 算法[J]. *计算机工程与设计*, 2014, 35(5): 1811-1815.
SHEN J J, JIANG H, WANG S. Improving RPCL algorithm based on cluster center counts and global distance vector [J]. *Computer Engineering and Design*, 2014, 35(5): 1811-1815. (in Chinese)
- [13] 谢娟英, 郭文娟, 谢维信, 等. 基于密度 RPCL 的 K-means 算法[J]. *西北大学学报(自然科学版)*, 2012, 42(4): 570-576.
XIE J Y, GUO W J, XIE W X, et al. A density RPCL based K-means algorithm [J]. *Journal of Northwest University (Natural Science Edition)*, 2012, 42(4): 570-576. (in Chinese)
- [14] 国家能源局. 风电场功率预测预报管理暂行办法[J]. *太阳能*, 2011(14): 6-7.
National Energy Administration. Interim management of wind power generation prediction [J]. *Solar Energy*, 2011(14): 6-7. (in Chinese)

(责任编辑 吕小红)