

L-M 优化 BP 算法在短期负荷预测中的应用

代小红¹ 王光利²

(重庆工商大学计算机科学与信息工程学院 重庆 400067)¹ (重庆邮电大学生物信息学院 重庆 400065)²

摘要 在分析传统 BP 算法的不足的基础上,提出了将 Levenberg-Marquardt 优化法与神经网络模型相结合的 L-M 优化 BP 算法。此方法与传统算法相比学习速度得到了提高,网络的收敛加快,尽量避免了系统陷入局部最小;针对某电力局某地区的单条线路的实际数据,采用基于 Levenberg-Marquardt 优化的 BP 算法的神经网络模型对其进行了仿真,结果表明该方法具有较高的预测精度和较强的适应能力。

关键词 短期负荷预测, L-M 优化法, BP 算法, 预测误差

中图分类号 TM71 文献标识码 A

Application of L-M Optimized BP Algorithm in Short-term Power Load Forecast

DAI Xiao-hong¹ WANG Guang-li²

(College of Computer and Information Engineering, Chongqing Technology and Business University, Chongqing 400067, China)¹

(School of Bioinformation, Chongqing University of Post and Telecommunications, Chongqing 400065, China)²

Abstract Analyzing the deficiency of the traditional BP algorithm, combining Levenberg-Marquardt optimized algorithm and a neural network forecasting method, this paper put forward a L-M optimized BP algorithm, which quickens the train, improves stability and avoids trapping into local minimum. For some area power supply load of Power Corporation in somewhere, a short-term load forecast was simulated based on L-M optimized BP algorithm. Analyzing the simulation results, it shows the L-M optimized BP algorithm has better forecast precision and adaptive capacity.

Keywords Short-term load forecasting, L-M optimized, BP algorithm, Forecast precision

1 概述

电力系统负荷预测是指一定条件下,采用一定方法处理系统的过去和未来负荷的数学方法,这些条件包括本系统的运行状态与特性、自然环境条件、系统增容的决策等因素。在满足一定系统预定要求的前提条件下,确定本系统在未来某一个特定时刻的负荷值。而电力系统负荷短期预测是指一年之内以月为时间单位的电力负荷预测,在某些情况下可以将预测时间单位划分到周(7天)、天(24小时)、小时。在通常情况下,电力负荷预测是指测算未来一月(30天)或未来一周(7天)甚至是未来某一天的负荷指标情况,也预测未来一天24小时中的负荷。负荷预测的核心问题是预测的数学模型。随着现代科学技术的快速发展,负荷预测技术的研究也在不断深化,各种各样的负荷预测方法不断涌现。

神经网络技术运用于电力系统的负荷预测,目前来说还是较为新型的研究方法。这种技术的优点是在一定的环境条件下,可以进行智能化的处理,仿人脑的思维方法;特别是它具备并行分析、自学习和任意逼近连续函数的特点,对一些复杂的非结构性、非精确性规律具有较强的自适应能力。其具备的自学习和自适应功能是常规算法和专家系统技术所不可比的。当前采用的基于 BP 算法的多层感知器网络运用广

泛,是电力系统普遍的负荷预测解决方案,在其他领域中得到了广泛的应用与推广。但是基于 BP 网络算法采用的是沿梯度下降的搜索求解方法,这就可能在系统中出现网络收敛速度慢、易陷入局部极小等现象。在一些研究中提出了各种方法进行改进,以解决此类问题现象,这些研究在不同程度上提高了学习速度,网络的收敛得到了加快的同时尽量避免了陷入局部极小。这种改进优化方法^[1]的算法源于数学优化算法。Levenberg-Marquardt(非线性阻尼最小二乘法)具有二阶的收敛速度,避免了计算海森矩阵。基于 Levenberg-Marquardt 优化法的中型神经网络(几百个权值)具有较快的训练函数方法。利用某电力局某地区的单条线路的实际数据进行仿真,获得了较好的预测效果。

2 L-M 优化法的 BP 模型学习算法

2.1 原理分析

Levenberg-Marquardt 法又称为非线性阻尼最小二乘法。若关于参量的数学模型是非线性的,用最小二乘法处理试验数据的曲线拟合问题就是非线性最小二乘问题。例如,拟合函数为

$$y = f(t; x_1, \dots, x_n) \quad (1)$$

式中, t 为自变量, $x_1, x_2, \dots, x_n$ 为参数, $f(t; x_1, \dots, x_n)$ 关于

参数为非线性。若用 $f(t; x_1, \dots, x_n)$ 拟合试验数据 (t_i, y_i) , $i=1, 2, \dots, m$, 则用最小二乘法确定参数时就得到关于

$$\varphi(x_1, x_2, \dots, x_n) = \sum_{i=1}^m [f(t_i; x_1, \dots, x_n) - y_i]^2, m > n \quad (2)$$

的极小问题。

上述问题可以归纳为以下形式。设 $f: D \subset R^n \rightarrow R^m$, $f = (f_1, \dots, f_m)^T$, 定义函数

$$\varphi(x) = \frac{1}{2} f(x)^T f(x) \quad (3)$$

显然 $\varphi: D \subset R^n \rightarrow R$, 于是非线性最小二乘问题就是求

$$\min_{x \in D} \varphi(x) = \min_{x \in D} \frac{1}{2} f(x)^T f(x) \quad (4)$$

以上是求多元函数 $\varphi(x)$ 的极小问题, 根据存在极值的条件, 如果 f 在 D 上可微, 则 $\varphi(x)$ 对 x 求导得到 $\varphi(x)$ 的极小, 满足方程, 有

$$g(x) = \nabla \varphi(x) = Df(x)^T f(x) = 0 \quad (5)$$

称其为法方程, 式(5)中有

$$Df(x)^T = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_2}{\partial x_1} & \dots & \frac{\partial f_m}{\partial x_1} \\ \vdots & \vdots & \dots & \vdots \\ \frac{\partial f_1}{\partial x_n} & \frac{\partial f_2}{\partial x_n} & \dots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

可把 $f(x)$ 在 x^k 点进行线性展开如下

$$f(x) \approx Df(x^k)(x - x^k) + f(x^k) = l_k(x) \quad (6)$$

在式(6)中用 $l_k(x)$ 替代 $f(x)$, $Df(x^k)$ 替代 $Df(x)$, 得到 $Df(x^k)^T l_k(x) = 0$

并将它的解记作 x^{k+1} , 于是有

$$x^{k+1} = x^k - [Df(x^k)^T Df(x^k)]^{-1} Df(x^k)^T f(x^k) \quad (7)$$

在此称此式为高斯-牛顿法, 运用式(6)中的记号, 并设

$$G(x) = Df(x)^T Df(x) \quad (8)$$

则式(7)可改写成

$$x^{k+1} = x^k - G(x^k)^{-1} g(x^k), k=0, 1, \dots \quad (9)$$

它在形式上与式(6)的牛顿法有些类似, 但它又不是牛顿法; 因为式(6)的牛顿法要求解 $f(x)$ 的 2 阶导数, 较为复杂, 而在此只需求解一阶导数 $Df(x^k)$, 这样就变得较简单, 且仍具有良好的性质。事实上, 若 $f(x) = Ax - b$, 其中 $A \in R^{m \times n}$, $n < m$, $\text{rank}(A) = n$, 则式(6)的 $\varphi(x)$ 是二次函数。这时有

$$g(x) = \nabla \varphi(x) = A^T(Ax - b), G(x) = A^T A$$

式中, $A^T A$ 为对称正定矩阵。对于任意初始近似 $x^0 \in R^n$, 用高斯-牛顿法经一步迭代则可得问题(5)的解 $x^1 = x^* = (A^T A)^{-1} A^T b$, 称其为二次终止性。对于一些常见的非线性函数 $f(x)$, 由式(8)定义的 $G(x)$ 知, 它是对称非负定矩阵。若 $G(x^k)$ 正定, 式(9)中记 $p_k = -G(x^k)^{-1} g(x^k) = -G(x^k)^{-1} \nabla \varphi(x^k)$, 则

$$p_k^T \nabla \varphi(x^k) = -p_k^T G(x^k) p_k < 0$$

其表明向量 p_k 和 $-\nabla \varphi(x^k)$ 的方向是一致的, 是 $\varphi(x)$ 在点 x^k 处的下降。为了避免 $G(x^k)$ 奇异或病态, 可以增加一个阻尼项, 即令

$$\hat{G}(x^k) = G(x^k) + \mu_k I$$

当阻尼因子 $\mu_k > 0$ 时, 显然 $\hat{G}(x^k)$ 是对称正定矩阵, 从而

$$p_k(\mu_k) = -\hat{G}(x^k)^{-1} \nabla \varphi(x^k)$$

是 $\varphi(x)$ 在点 x^k 处的下降方向。于是可构造迭代法

$$x^{k+1} = x^k - \hat{G}(x^k)^{-1} \nabla \varphi(x^k) \quad k=0, 1, \dots$$

称其为阻尼最小二乘法, 或称为 L-M 法, 它是高斯-牛顿法的改进。当 $\mu_k > 0$ 时总可保证 $\varphi(x^{k+1}) < \varphi(x^k)$, 因此 $\{x^k\}$ 总是收敛的; 但 $\mu_k > 0$ 太大, 序列 $\{x^k\}$ 收敛速度会下降, 而 μ_k 太小其收敛域又过小, 初始近似 x^0 受限制。当然要想取一个较合适的 μ_k 是困难的, 但是原则上 μ_k 是一个小的正数, 例如 $\mu_k = 10^{-4} \sim 10^{-2}$ 。通常取 $\mu_k = 10^{-2}$, 然后通过计算加以限制。

2.2 L-M 优化的 BP 算法模型

此算法采用 3 层的 BP 算法模型。隐层神经元节点的输出为

$$y_i = f(\sum_j w_{ij} x_j - \theta_i) = f(\text{net}_i)$$

式中, $\text{net}_i = \sum_j w_{ij} x_j - \theta_i$ 。

输出节点的计算输出为

$$O_l = f(\sum_i T_{li} y_i - \theta_l) = f(\text{net}_l)$$

式中, $\text{net}_l = \sum_i T_{li} y_i - \theta_l$ 。

输出层节点的误差为

$$E = \frac{1}{2} \sum_l (t_l - O_l)^2 = \frac{1}{2} \sum_l (t_l - f(\sum_i T_{li} y_i - \theta_l))^2$$

$$= \frac{1}{2} \sum_l (t_l - f(\sum_i T_{li} f(\sum_j w_{ij} x_j - \theta_i) - \theta_l))^2$$

则输出层节点权值的梯度 $\frac{\partial E}{\partial T_{li}} = \sum_{k=1}^m \frac{\partial E}{\partial O_k} \frac{\partial O_k}{\partial T_{li}} = \frac{\partial E}{\partial O_l} \frac{\partial O_l}{\partial T_{li}}$, E 是多个 O_k 的函数, 但只有一个 O_l 与 T_{li} 有关, 各 O_k 间相互独立。故

$$\frac{\partial E}{\partial O_l} = \frac{1}{2} \sum_k -2(t_k - O_k) \cdot \frac{\partial O_k}{\partial O_l} = -(t_l - O_l)$$

又因为

$$\frac{\partial O_l}{\partial T_{li}} = \frac{\partial O_l}{\partial \text{net}_l} \frac{\partial \text{net}_l}{\partial T_{li}} = f'(\text{net}_l) \cdot y_i$$

令 $F = -(t_l - O_l)$, 则

$$\frac{\partial E}{\partial T_{li}} = F \cdot f'(\text{net}_l) \cdot y_i$$

$$\frac{\partial F}{\partial T_{li}} = \frac{\partial O_l}{\partial \text{net}_l} \frac{\partial \text{net}_l}{\partial T_{li}} = f'(\text{net}_l) \cdot y_i$$

$$G(T_{li}) = \frac{\partial F}{\partial T_{li}} * \left(\frac{\partial F}{\partial T_{li}} \right)^T = (f'(\text{net}_l) \cdot y_i) * (f'(\text{net}_l) \cdot y_i)'$$

$$p(T_{li}) = -[G(T_{li}) + u(T_{li})I]^{-1} * \frac{\partial E}{\partial T_{li}}$$

隐层神经元节点的梯度 $\frac{\partial E}{\partial w_{ij}} = \sum_l \sum_i \frac{\partial E}{\partial O_l} \frac{\partial O_l}{\partial y_i} \frac{\partial y_i}{\partial w_{ij}}$, E 是基于多个 O_l 的函数, 就某一个 w_{ij} , 对应于一个 y_i , 它与所有的 O_l 相关, 其中

$$\frac{\partial E}{\partial O_l} = \frac{1}{2} \sum_k -2(t_k - O_k) \cdot \frac{\partial O_k}{\partial O_l} = -(t_l - O_l)$$

$$\frac{\partial O_l}{\partial y_i} = \frac{\partial O_l}{\partial \text{net}_l} \frac{\partial \text{net}_l}{\partial y_i} = f'(\text{net}_l) \cdot \frac{\partial \text{net}_l}{\partial y_i} = f'(\text{net}_l) \cdot T_{li}$$

$$\frac{\partial y_i}{\partial w_{ij}} = \frac{\partial y_i}{\partial net_i} \frac{\partial net_i}{\partial w_{ij}} = f'(net_i) \cdot x_j$$

$$\text{则} \frac{\partial E}{\partial w_{ij}} = - \sum_l (t_l - O_l) \cdot f'(net_l) \cdot T_{li} \cdot f'(net_i) \cdot x_j$$

令 $F = -(t_l - O_l)$, 则

$$\frac{\partial F}{\partial w_{ij}} = \frac{\partial O_l}{\partial net_l} \frac{\partial net_l}{\partial y_i} \frac{\partial y_i}{\partial w_{ij}} = f'(net_l) \cdot T_{li} \cdot f'(net_i) \cdot x_j$$

$$G(w_{ij}) = \frac{\partial F}{\partial w_{ij}} * \left(\frac{\partial F}{\partial w_{ij}} \right)^T$$

$$p(w_{ij}) = -[G(w_{ij}) + u(w_{ij})I]^{-1} * \frac{\partial E}{\partial w_{ij}}$$

则输出层、隐层神经元节点的权值修正公式为

$$T_{li}(k+1) = T_{li}(k) + p(T_{li}) \quad (10)$$

$$w_{ij}(k+1) = w_{ij}(k) + p(w_{ij}) \quad (11)$$

同理,输出层和隐层神经元节点的阈值修正公式为

$$\theta_l(k+1) = \theta_l(k) + p(\theta_l) \quad (12)$$

$$\theta_i(k+1) = \theta_i(k) + p(\theta_i) \quad (13)$$

传递函数 $f(x)$ 的导数公式及函数 $f(x) = \frac{1}{1+e^{-x}}$ 存在关系

$$f'(x) = f(x) \cdot (1-f(x)), \text{则}$$

$$f'(net_l) = O_l \cdot (1-f(O_l))$$

$$f'(net_i) = y_i \cdot (1-f(y_i))$$

3 L-M 优化法的 BP 预测模型仿真分析

在此利用基于 L-M 优化法的 BP 算法,进行某电力局某地区一条线路的供电负荷为期一天(24h)的短期负荷预测,采集单条线路从 2008 年 2 月 17 日到 2008 年 3 月 24 日的 24 点电力负荷数据,由这些数据可构造出以时间单位为序的 28 个一天(24 h)数据样本。顺序取出前 21 个数据样本训练神经网络的权值和阈值,称为训练样本;后面 7 个数据样本评价神经网络的性能,称为评价样本。利用 3 层基于 L-M 优化法的 BP 算法进行训练,对 21 个训练样本采用批量处理方式,即仅仅当所有的输入数据样本都被提交以后,网络权重和阈值才被更新。通过仿真,11 步训练误差就会降至目标误差曲线 10^{-5} 以下,由此可见基于 L-M 优化法的 BP 模型的收敛速度快,且避免了陷入局部极小等。

图 1 示出 2008 年 2 月 25 日至 2008 年 3 月 17 日共计 21 个训练样本的 4 时刻样本值与预测值曲线,在此经过 11 步训练后,基于 L-M 法的 BP 模型算法已经达到了对训练样本的完全逼近。从图 2 中 7 个评价样本的 4 时刻样本值与预测值曲线可以看出,与传统 BP 算法模型相比,L-M 算法的电力负荷预测曲线具有与样本曲线更为接近、预测误差较小等优点。

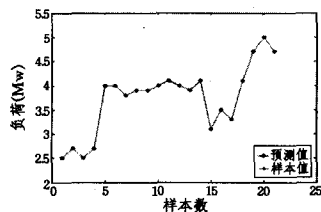


图 1 2/25~3/17 4 时刻样本值与预测值曲线

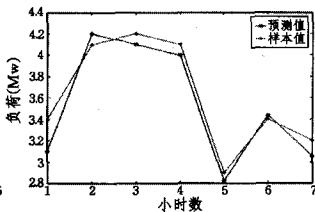


图 2 3/18~3/24 共 7 个评价样本 4 时刻值与预测值曲线

由 2008 年 3 月 18 日到 2008 年 3 月 24 日共计 1 周(7 天)的预测值曲线和样本值曲线(图 3—图 4)可以看出,该算法的预测精度比传统 BP 算法的预测精度得到了明显的提高,且峰值误差得到了一定的改善,但在某个别峰值处,误差较传统算法变得尖锐,不那么平滑。这就更加验证了 L-M 算法在预测峰值误差方面比传统 BP 算法具有更好的逼近。

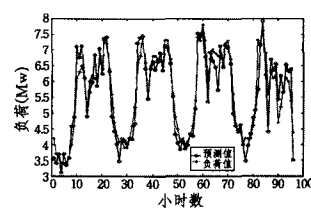


图 3 3/18~3/21(24h)样本值与预测值曲线

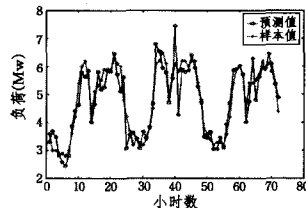


图 4 3/22~3/24(24h)样本值与预测值曲线

结束语 利用基于 L-M 优化法和神经网络相结合的算法模型,对某电力局某地区的单条线路的实际数据进行仿真。由仿真结果可知,此预测算法模型较传统 BP 神经网络模型有了较大改善,提高了系统的学习速度,大大加快了网络的收敛,从而避免了陷入局部极小等现象,预测结果更稳定,预测精度更高。对负荷波动较大的单线路负荷预测具有较高的实际应用价值。

参考文献

- [1] 赖晓平. 混合模型神经网络在短期负荷预测中的应用[J]. 控制理论与应用, 2000(1): 69-72
- [2] 吴祈宗. 运筹学与最优化方法[M]. 北京: 机械工业出版社, 2003
- [3] 罗亚中, 唐国金, 田蕾. 基于模拟退火算法的最优控制问题全局优化[J]. 南京理工大学学报, 2005, 29(2): 144-148
- [4] 马光文, 王黎. 水电竞价上网优化运行[M]. 成都: 四川科学技术出版社, 2003
- [5] 招海丹, 吴捷, 杨苹, 等. 一个综合智能化电力短期负荷预测系统的研究[J]. 电网技术, 2000(12): 45-48
- [6] 康重庆, 夏清, 沈瑜, 等. 电力系统负荷预测的综合模型[J]. 清华大学学报: 自然科学版, 1999(1): 8-11
- [7] 赖晓平, 周鸿兴, 田发中. 电力系统短期负荷预测的混合模型神经网络方法[J]. 电网技术, 2000(1): 47-51
- [8] 胥红星, 舒永录, 原冠秀. 一个混沌系统最终有界集及其在同步中的应用[J]. 重庆工商大学学报: 自然科学版, 2008, 25(6): 564-568
- [9] 王士同. 神经模糊系统及其应用[M]. 北京: 北京航空航天大学出版社, 1998
- [10] 林雄. 神经网络在电网负荷预测上的应用研究[J]. 信息技术, 2008(9): 111-113, 143
- [11] 秦海超, 王玮, 周晖, 等. 人体舒适度指数在短期电力负荷预测中的应用[J]. 电力系统及其自动化学报, 2006, 18(2): 63-66
- [12] 黄勤, 龚海清, 刘金亨, 等. 基于改进的遗传神经网络入侵检测系统[J]. 重庆理工大学学报: 自然科学版, 2010, 24(2): 83-86