

基于BP网的中长期负荷预测因素优化选择

朱继萍, 戴君

(西安文理学院机械电子工程系, 西安 710065)

摘要: 基于神经网络原理, 设计一个由输入层、隐含层和输出层组成的三层BP网络模型, 利用神经网络高度非线性建模能力, 选取影响电力负荷的一些经济因素作为BP神经网络的输入变量, 采用新定义的方差贡献法对输入变量进行优化选择, 对预测精度的影响进行探讨。仿真结果证明, 采用方差贡献法对影响中长期电力负荷预测的相关因素进行优化选择是可行有效的。

关键词: BP网络; 中长期负荷预测; 方差贡献法; 优化选择

Optimization Selection of Correlative Factors for Medium and Long Term Load Forecasting Based on BP Network

ZHU Ji-ping, DAI Jun

(Department of Mechanical and Electronic Engineering, Xi'an University of Arts and Science, Xi'an 710065)

【Abstract】 Based on the theory of Artificial Neural Network(ANN), a three-layer Back Propagation(BP) network is proposed. The idea is to forecast medium and long term load by using the ability of ANN to nonlinear system factors. Some economic factors are selected as inputs for the BP ANN model. Variance contribution method is used for the optimization selection of correlative factors, and forecasting accuracy is discussed. Simulation results show that the method is feasible.

【Key words】 Back Propagation(BP) network; medium and long term load forecasting; variance contribution method; optimization selection

1 概述

负荷预测是城网规划的基础, 对规划的质量起关键作用^[1]。由于城市建设发展的过程是变化的, 因此经济预测数据及经济的增长也是变化的, 未来发展的不确定性影响着负荷预测的准确性。尽管目前已有多种中长期负荷预测方法, 新的预测方法仍在不断研究^[2]。预测方法, 但各种方法都有局限性, 无法保证负荷预测的绝对准确。

为了提高负荷预测的准确性, 在进行负荷预测时总希望将影响负荷的所有因素都作为输入变量, 特别是那些作用明显的因素。但是如果选择的相关因素太多, 意味着在资料收集上要全面准确, 而在实际中有些资料的收集是很困难的, 并且用神经网络(Artificial Neural Network, ANN)进行预测时, 选择的输入变量越多, 预测的模型就越复杂, 预测的精度也难以保证。从这点考虑, 一般不希望选择的因素太多, 尤其是对负荷影响较小的因素尽量不予选择, 而尽量选择那些对负荷有明显影响的相关因素作为预测模型的输入变量。

2 神经网络的结构^[3-4]

因为中长期负荷预测的非线性程度较高, 所以采用三层网络。选用有一个隐含层的BP(Back Propagation)网络进行中长期负荷预测。图1是本文用于预测的BP网络模型。

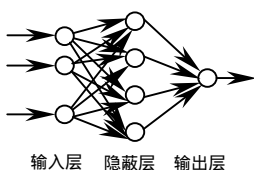


图1 BP神经网络的预测模型

BP网络是一种非线性映射的多层前馈高级人工神经元

网络, 理论上可以映射任意复杂的非线性关系。取一个隐含层, 输入层为对负荷有影响的一些相关因素(如GDP和工农业生产总值), 输出为待预测年的负荷, 节点激活函数取Sigmoid函数。通过代入历史数据进行神经网络训练, 用训练好的网络进行负荷的预测。

3 相关影响因素的优化选择

3.1 方差贡献的定义

在用BP网络进行中长期负荷预测时, 总是选取与负荷密切相关的一些影响因素(如重工业生产总值、农业生产总值)作为网络模型的输入变量。但是, 这些经济因素在不同地区对负荷的影响是不同的, 有些地区重工业生产总值是对负荷最有影响的输入变量, 而有些地区可能是第三产业对负荷变化影响最大。为了保证预测的精度, 在建立模型时, 必须对那些对负荷有影响的因素进行优化选择, 以便选择出对负荷有重要影响的因素作为模型的输入变量, 达到既简化预测模型又保证预测精度的目的。

方差贡献法用于分析各类影响因素对负荷的影响程度。假设输入变量为 x_1, x_2, \dots, x_n , 负荷的预测值为 \hat{y} , 真实值为 y , 可以根据式(1)

$$Q_e = \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (1)$$

计算出残差平方和 Q_e 。

基金项目: 西安文理学院自然科学专项科研基金资助项目(KY200530)

作者简介: 朱继萍(1975-), 女, 助教, 主研方向: 信息处理, 神经网络应用; 戴君, 副教授

收稿日期: 2007-10-20 **E-mail:** kuihua577@yahoo.com.cn

如果在 x_1, x_2, \dots, x_n 中去除一个变量 X_j 后再重新训练 BP 网络进行负荷预测, 那么预测结果的残差平方和 Q_e 将变化。设去除变量 X_j 后残差平方和为 $Q_e(-x_j)$, 定义 $Q(-x_j)$ 为去除 X_j 的方差贡献, 即

$$Q(-x_j) = Q_e(-x_j) - Q_e(0) \quad (2)$$

其中, $Q_e(0)$ 为未去除 X_j 时的残差平方和。

$Q(-x_j)$ 的大小反映了被去除的相关因素 X_j 对负荷预测的贡献大小。

类似地, 引入变量 X_k 后的方差贡献 $Q(+x_k)$ 为

$$Q(+x_k) = Q_e(0) - Q_e(+x_k) \quad (3)$$

其中, $Q_e(0)$ 为未引入变量 X_k 时的残差平方和。

在进行输入变量的选择时, 一方面先从一个变量开始, 按照各自对负荷的影响程度, 从大到小依次作为神经网络的输入变量; 另一方面当后引入的相关因素变得对负荷影响不明显时, 先引入的相关因素不予保留, 随时将它从输入变量中去除。在对输入变量进行选择时, 有时是引入相关因素, 有时是去除相关因素, 并且每一步都要进行 F 检验, 以保证每次在引入新的相关因素之前, 输入变量中只包含对负荷有明显影响的相关因素, 直到所有对负荷有明显影响的相关因素都作为网络的输入变量。

3.2 相关因素的 F 检验

要确定被选择的相关因素能否作为输入变量必须进行检验, 否则选择将毫无意义。因此, 需要给出一个标准值, 对引入的相关因素的方差贡献进行 F 检验^[5-6]。假设已经选取了 l 个相关因素, 为了方便, 认为这 l 个相关因素为 x_1, x_2, \dots, x_l , 把它们作为输入变量时负荷预测对应的残差平方和为 $Q_e(0)$ 。如前所述, 在选择下一个相关因素时, 除了这已选的 l 个相关因素以外, 应该在剩下的所有相关因素中选择一个作为输入变量, 重新进行预测后使新的残差平方和最小, 这个相关因素就是 x_{l+1} , 对应的残差平方和为 $Q_e(+x_{l+1})$, 这时的方差贡献为

$$Q(+x_{l+1}) = Q_e(0) - Q_e(+x_{l+1}) \quad (4)$$

增加输入变量 x_{l+1} 后的方差贡献 $Q(+x_{l+1})$ 越大, 说明新的输入变量 x_{l+1} 对负荷的影响作用越大。又由于方差贡献 $Q(+x_{l+1})$ 不仅与输入变量 x_{l+1} 有关, 而且与 $(x_1, x_2, \dots, x_l, x_{l+1})$ 的整体有关, 因此用 x_{l+1} 的方差贡献 $Q(+x_{l+1})$ 与 x_{l+1} 作为输入变量时的残差平方和 $Q_e(+x_{l+1})$ 的比值 $Q(+x_{l+1})/Q_e(+x_{l+1})$ 来衡量 x_{l+1} 对负荷所起作用的大小更为恰当。

设统计量

$$F_{(+x_{l+1})} = \frac{Q(+x_{l+1})/(n-1-l-[(n-1)-(l+1)])}{Q_e(+x_{l+1})/n-1-(l+1)} = \frac{Q(+x_{l+1})}{Q_e(+x_{l+1})/n-l-2} = \frac{Q(+x_{l+1})}{Q_e(+x_{l+1})} (n-l-2) \quad (5)$$

在一定显著水平 α 下, 由于比值 $Q(+x_{l+1})/Q_e(+x_{l+1})$ 越小, 说明相关因素 x_{l+1} 对负荷的作用越小, 因此确定临界值 $F_{\alpha}(1, n-l-2)$ 为:

(1) 当 $F_{(+x_{l+1})} > F_{\alpha}(1, n-l-2)$ 时, 相关因素 x_{l+1} 对负荷的影响作用明显, 要把 x_{l+1} 增加为输入变量;

(2) 当 $F_{(+x_{l+1})} < F_{\alpha}(1, n-l-2)$ 时, 相关因素 x_{l+1} 对负荷的影响作用不明显, 不能把 x_{l+1} 增加为输入变量。

那么挑选新的输入变量的工作就结束了。

3.3 相关因素的去除

应该选择方差贡献较大的相关因素, 作为输入变量, 但刚开始被选择作为输入变量的相关因素可能会因为后来增加的作为输入变量的相关因素而变得对负荷的影响作用不明显。因此, 每选一个新的相关因素作为输入变量之前, 要检查已选择的输入变量中有没有对负荷作用不明显的输入变量, 这种输入变量要及时去除。

假定已经选择了 l 个相关因素 x_1, x_2, \dots, x_l 作为输入变量, 此时负荷预测对应的残差为 $Q_e(0)$, 在引入新的相关因素之前, 要对已有输入变量中各个变量对负荷作用的显著性加以检验。为此, 要在已选的 l 个相关因素中研究单个相关因素 X_j ($j=2, 3, \dots, l$) 对负荷的作用。

若去除相关因素 X_j , 那么剩余的 $(l-1)$ 个相关因素再作为输入变量时, 进行负荷预测相应的残差平方和为 $Q_e(-x_j)$, 则 X_j 的方差贡献为

$$Q(-x_j) = Q_e(-x_j) - Q_e(0) \quad (6)$$

$Q(-x_j)$ 越大, 说明 X_j 对负荷的作用越大。

用被去除 X_j 的方差贡献 $Q(-x_j)$ 与包括 X_j 作为输入变量时负荷预测的残差平方和 $Q_e(0)$ 的比值 $Q(-x_j)/Q_e(0)$ 来衡量被去除的 X_j 对负荷的作用大小。

由于 $Q_e(0)$ 与 $j=1, 2, \dots, l$ 无关, 因此在输入变量中要去除的相关因素将是其方差贡献最小的一个, 记为 $Q(-x_k)$, 即

$$Q(-x_k) = \min_{j=1, 2, \dots, l} Q(-x_j) \quad (7)$$

4 仿真预测

以陕西省 1990 年 - 2000 年的历史数据作为训练样本, 以 2001 年和 2002 年作为检验样本, 构造了一个三层的多输入单输出 BP 神经网络, 对陕西算例用 BP 人工神经网络进行预测。选择国内生产总值 (X_1)、第一产业产值 (X_2)、第二产业产值 (X_3)、第三产业产值 (X_4)、重工业总产值 (X_5)、轻工业总产值 (X_6)、农业总产值 (X_7) 这 7 个相关因素作为输入变量, 用方差贡献法分析不同相关因素的组合对负荷预测结果的影响, 见表 1。

表 1 去除各影响因素后的方差贡献

影响因素	方差贡献 Q_j	F 值
各种因素有	-	-
去除 X_1	-320.210 3	5.352 1
去除 X_2	-112.809 2	5.124 1
去除 X_3	172.065 5	4.810 8
去除 X_4	-417.541 1	5.459 2
去除 X_5	-3.355 3e+ 003	8.689 9
去除 X_6	-768.796 3	5.845 5
去除 X_7	-779.437 7	5.857 5

从表 1 中可以看出, 根据 F 值, 查《 F 检验临界值表》确定 $F_{0.05}(l, n-l-1)$, l 为自变量个数, n 为样本个数, 对两者进行比较, 若 $F > F_{0.05}(l, n-l-1)$, 说明 X_k 对 y 的贡献大, 应不予剔除; 若 $F < F_{0.05}(l, n-l-1)$, 则剔除。本算例中 $l=7$, $n=13$, 查表之后 $F_{0.05}(7, 5)=4.88$, 比较后可见, 变量 X_3 的贡献最小, 在选择输入变量时可以去除; X_5 对应的 F 值最大, 即这个变量贡献最大, 在预测时应该被保留作为输入变量。通过分析单个输入变量的方差贡献后, 选择变量 X_5 为必选变量, 去除变量 X_3 。在去除变量 X_3 并且保留变量 X_5 的同时, 再依次去除其余的变量, 分析同时去除 2 个变量对负荷影响的贡献大小, 结果如表 2 所示。 (下转第 230 页)