

基于函数逼近的多层前馈神经网络灵敏度分析

吴跃波, 杨景曙

WU Yue-bo, YANG Jing-shu

解放军电子工程学院 702 室, 合肥 230037

Department of 702, PLA Electronic Engineering Institute, Hefei 230037, China

E-mail: bobi--2002@163.com

WU Yue-bo, YANG Jing-shu. Sensitivity analysis of multi-layer feed-forward neural networks based on approximating functions. *Computer Engineering and Applications*, 2010, 46(5): 36-39.

Abstract: Sensitivity analysis is vital in the design of neural networks. The existing approaches to the sensitivity analysis impose some limitations on network input and weight perturbations or can't apply exact enough result. This paper gives a better sensitivity formula of multi-layer feed-forward neural networks by using two approximating functions based on Piché's stochastic model. It has the form convenient for the computation of sensitivity without introducing addition limitations, and brings less error than other methods. The result of computer simulation proves this formula is correct and exact.

Key words: neural networks; sensitivity; approximating function

摘要: 神经网络灵敏度分析对网络结构设计、硬件实现等具有重要的指导意义, 已有的灵敏度计算公式对权值和输入扰动有一定限制或者计算误差较大。基于 Piché 的随机模型, 通过使用两个逼近函数对神经网络一类 Sigmoid 激活函数进行高精度逼近, 获得了新的神经网络灵敏度计算公式, 公式取消了对权值扰动和输入扰动的限制, 与其他方法相比提高了计算精度, 实验证明了公式的正确性和精确性。

关键词: 神经网络; 灵敏度; 逼近函数

DOI: 10.3778/j.issn.1002-8331.2010.05.012 文章编号: 1002-8331(2010)05-0036-04 文献标识码: A 中图分类号: TP18

1 引言

神经网络灵敏度研究的是当网络参数有一定扰动时, 网络输出扰动的变化情况, 这对于网络的结构优化^[1-2]、硬件实现^[3-4]、学习算法的选择^[5]和分类及泛化能力^[6]的研究具有指导性的作用。目前应用最为广泛的是具有 Sigmoid 型激活函数的多层前馈神经网络, 研究正是基于这种网络。

对于具有连续可导激活函数的多层前馈神经网络, 灵敏度研究目前主要有两种^[7]: 偏导数灵敏度 (Partial Derivate Sensitivity) 和随机灵敏度 (Stochastic Sensitivity)。近来对灵敏度分析理论的研究多偏重于随机灵敏度^[8-10]。随机灵敏度的基本理论模型由 Piché 提出^[3], 模型的优点是不必关心随机变量的概率分布, 只需知道变量的一些数字特征, 然而计算方法却有着一定局限性: 对 Sigmoid 型激活函数的网络不能获得一个统一的解析表达式; 只对输入和权值的扰动充分小的时候适用, 这大大限制了其适用范围。针对以上不足, Daniel 和 Xuequan 构造了一类函数来逼近 Sigmoid 型激活函数^[8], 获得了灵敏度的统一解析表达式, 同时取消了对输入和权值的扰动充分小的限制。然而计算对比和仿真实验表明, 在一些情况下这种逼近精度是不够的, 甚至会带来很大的计算误差, 不利于实际中的应用, 为此提出了一个新的更为精确的逼近函数。

先给出随机模型及灵敏度的定义, 之后提出了新的激活函数逼近公式, 接着基于 Piché 的随机模型得到了更精确的灵敏度公式, 同样也取消了对要求输入和权值的扰动充分小的限制, 最后将其和文献^[8]的结果进行对比, 通过实验验证了方法的正确性和准确性。

2 随机模型及灵敏度定义

工作基于 Piché 的随机模型, 介绍这一模型及灵敏度的定义。多层前馈神经网络参数说明如下: 网络各层用 0 到 L 来表示, 第 0 层为输入层, 第 L 层为输出层; 第 l 层的节点个数用 N_l 表示; $x_i^l, \Delta x_i^l$ 为第 l 层第 i 个节点的输入和输入扰动; $w_{ji}^l, \Delta w_{ji}^l$ 为第 $l-1$ 层第 i 个节点与第 l 层第 j 个节点的权值和权值扰动; y_i^l 为输入和权值无扰动时第 l 层第 i 个神经元的输出; Δy_i^l 为由输入和权值扰动造成的输出误差; 激活函数用 $g(\cdot)$ 表示。

Piché 随机模型描述如下:

(1) 各层各个节点的输入 x_i^l 为独立同分布的随机变量, 均值 $\mu_x=0$, 方差为 σ_x^2 ;

(2) 各层各个权值 w_{ji}^l 为独立同分布的随机变量, 均值 $\mu_w=$

作者简介: 吴跃波 (1984-), 男, 在读硕士研究生, 研究兴趣为神经网络、雷达信号处理等; 杨景曙 (1950-), 男, 教授, 博士生导师, 研究领域为电子对抗, 模式识别, 神经网络, 检测理论及应用等。

收稿日期: 2008-09-10 修回日期: 2008-12-08

0,方差为 $\sigma_{w_i}^2$;

(3)输入扰动 Δx_i^l 为独立同分布的随机变量,均值 $\mu_{\Delta x_i^l}=0$,

方差为 $\sigma_{\Delta x_i^l}^2$;

(4)权值扰动 Δw_{ji}^l 为独立同分布的随机变量,均值 $\mu_{\Delta w_{ji}^l}=0$,

方差为 $\sigma_{\Delta w_{ji}^l}^2$;

(5)每个 $x_i^l, w_{ji}^l, \Delta x_i^l, \Delta w_{ji}^l$ 之间都是相互独立的。

根据这一模型,对于每一个神经元,由输入和权值扰动造成的输出扰动的均值为 0,所以输出扰动的方差反映了输出扰动的变化范围,因此使用输出扰动的方差与输出方差的比值定义灵敏度,即类似噪信比。整个网络的灵敏定义基于单个神经元的灵敏度,所以先给出单个神经元灵敏度的定义再给出整个网络的灵敏度。

定义 1 假定 X 与 ΔX 为神经元输入向量, W 与 ΔW 为权值和取值扰动矩阵,则该神经元的灵敏度为:

$$S = \sqrt{\frac{D(g((W+\Delta W) \cdot (X+\Delta X)) - g(W \cdot X))}{D(g(W \cdot X))}} \quad (1)$$

其中 $g(\cdot)$ 为神经元激活函数, $D(\cdot)$ 表示求方差。

定义 2 整个神经网络的灵敏度为网络输出层神经元的灵敏度。

3 新的逼近函数

神经网络的激活函数经常采用 Sigmoid 型的函数,考虑最常见的型如 $g(x) = (1 - e^{-\lambda x}) / (1 + e^{-\lambda x})$ ($\lambda > 0$, 为倾斜因子) 的激活函数。文献[8]从避免缩小适用范围、公式推导方便和获得统一解析表达三个角度考虑,获得了如下的逼近函数:

$$g_1(x) = \begin{cases} 1 - e^{-0.446\lambda^2 x^2} & x \geq 0 \\ -1 + e^{-0.446\lambda^2 x^2} & x < 0 \end{cases} \quad (2)$$

它在 $g_1(0), g_1(-\infty), g_1(+\infty)$ 精确反映了被逼近函数。从泰勒级数逼近的角度考察该逼近函数,它是一个偶数次幂的逼近,因此一个改进方向是添加奇数次幂来获得更高逼近精度,定义如下新的逼近函数:

$$g_{w1}(x) = \begin{cases} 1 - e^{-A_1 x^2} - B_1 x e^{-C_1 x^2} & x \geq 0 \\ -1 + e^{-A_1 x^2} - B_1 x e^{-C_1 x^2} & x < 0 \end{cases} \quad (3)$$

其中 A_1, B_1, C_1 为待定参数。

可见新函数在 $g_{w1}(0), g_{w1}(-\infty), g_{w1}(+\infty)$ 也精确反映了被逼近函数。求 A_1, B_1, C_1 时,归结为使逼近误差最小的优化问题:

$$\min_{A_1, B_1, C_1} e = \int_{-\infty}^{+\infty} (g(x) - g_{w1}(x))^2 dx \quad (4)$$

获得 A_1, B_1, C_1 如下:

$$\begin{aligned} A_1 &= 0.198\lambda^2 \\ B_1 &= -0.392\lambda \\ C_1 &= 0.320\lambda^2 \end{aligned} \quad (5)$$

该函数用于保证计算灵敏度表达式分母 $D(g(W \cdot X))$ 的高精度。但遗憾的是用它计算分子时仍不能获得一个解析表达式,为此,构造第二个逼近函数:

$$g_{w2}(x) = \begin{cases} 1 - B_2 e^{-A_2 x^2} - (1 - B_2) e^{-C_2 x^2} & x \geq 0 \\ -1 + B_2 e^{-A_2 x^2} + (1 - B_2) e^{-C_2 x^2} & x < 0 \end{cases} \quad (6)$$

获得 A_2, B_2, C_2 如下:

$$\begin{aligned} A_2 &= 3.041\lambda^2 \\ B_2 &= 0.380 \\ C_2 &= 0.220\lambda^2 \end{aligned} \quad (7)$$

该函数用于保证计算灵敏度表达式分子的更高精度。

新函数 $g_{w1}(x)$ 逼近误差约是 $g_1(x)$ 逼近误差的 1/80, $g_{w2}(x)$ 逼近误差也约达 $g_1(x)$ 逼近误差的 1/16。下图是 $\lambda=1$ 时, $g(x), g_1(x), g_{w1}(x), g_{w2}(x)$ 的对比图(由于都是奇函数,故只画出横坐标正部分)。

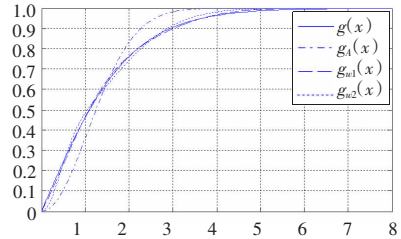


图1 $g(x)$ 及其逼近函数

可见, $g_{w1}(x), g_{w2}(x)$ 对 $g(x)$ 的逼近效果远比 $g_1(x)$ 好。高精度逼近函数是获得高精度灵敏度公式的保证。

4 灵敏度计算公式

利用前述的两个逼近函数来获得灵敏度的解析表达式。根据多层前馈神经网络的特点,在计算灵敏度时应采用自底至上(bottom-up)的方法。首先讨论单个神经元的情况,设第 l 层第 i 个神经元的输出为:

$$y_j^l = g\left(\sum_{k=1}^{N_{l-1}} w_{jk}^l x_k^l\right) \quad (8)$$

由输入和权值扰动造成的输出误差为:

$$\Delta y_j^l = g\left(\sum_{k=1}^{N_{l-1}} (w_{jk}^l + \Delta w_{jk}^l)(x_k^l + \Delta x_k^l)\right) - g\left(\sum_{k=1}^{N_{l-1}} w_{jk}^l x_k^l\right) \quad (9)$$

定义

$$\begin{aligned} \alpha &= \sqrt{N_{l-1}} \sigma_{x_i} \sigma_{w_j^l} \\ \beta &= \sqrt{\frac{\sigma_{\Delta x_i^l}^2}{\sigma_{x_i^l}^2} + \frac{\sigma_{\Delta w_{jk}^l}^2}{\sigma_{w_{jk}^l}^2} + \frac{\sigma_{\Delta x_i^l}^2}{\sigma_{x_i^l}^2} \cdot \frac{\sigma_{\Delta w_{jk}^l}^2}{\sigma_{w_{jk}^l}^2}} \end{aligned} \quad (10)$$

$$s = \frac{1}{\alpha} \sum_{k=1}^{N_{l-1}} w_{jk}^l x_k^l$$

$$t = \frac{1}{\alpha\beta} \sum_{k=1}^{N_{l-1}} (\Delta w_{jk}^l x_k^l + w_{jk}^l \Delta x_k^l + \Delta w_{jk}^l \Delta x_k^l)$$

根据 Piché 的随机模型,当 N_{l-1} 较大时(在实际应用中大于 10)即可认为 s, t 都服从标准正态分布,且相互独立。 s 和 t 的联合概率密度函数为:

$$f(s, t) = \frac{1}{2\pi} e^{-\frac{s^2 + t^2}{2}} \quad (11)$$

将式(10)代入到式(8)和(9),可得:

$$y_j^l = g(\alpha s) \quad (12)$$

$$\Delta y_j^l = g(\alpha s + \alpha\beta t) - g(\alpha s) \quad (13)$$

首先计算 y_j^l 的期望和方差:

$$E(y_j^l) = g(\alpha \cdot E(s)) = g(0) = 0 \quad (14)$$

$$D(y_j^l) = \int_{-\infty}^{+\infty} g^2(\alpha s) \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} ds \quad (15)$$

式(15)是无法获得解析表达式的,故用 $g_{a1}(x)$ 代替 $g(x)$ 来近似求解方差:

$$D(y_j^l) \approx \int_{-\infty}^{+\infty} (1 - e^{-A_1 \alpha^2 s^2} - B_1 \alpha s e^{-C_1 \alpha^2 s^2})^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} ds =$$

$$1 - \frac{2}{\sqrt{2A_1 \alpha^2 + 1}} + \frac{1}{\sqrt{4A_1 \alpha^2 + 1}} - \frac{\sqrt{2} B_1 \alpha}{\sqrt{\pi} (C_1 \alpha^2 + 0.5)} +$$

$$\frac{\sqrt{2} B_1 \alpha}{\sqrt{\pi} (A_1 \alpha^2 + C_1 \alpha^2 + 0.5)} + \frac{B_1^2 \alpha^2}{(4C_1 \alpha^2 + 1)^{\frac{3}{2}}} \quad (16)$$

其次计算 Δy_j^l 的期望和方差:

$$E(\Delta y_j^l) = 0 \quad (17)$$

$$D(\Delta y_j^l) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (g(\alpha s + \alpha \beta t) - g(\alpha s))^2 \frac{1}{2\pi} e^{-\frac{s^2+t^2}{2}} ds dt \quad (18)$$

式(18)也是无法获得解析表达式的,故用 $g_{a2}(x)$ 代替 $g(x)$ 来近似求解方差,得到:

$$D(\Delta y_j^l) \approx \frac{B_2^2}{\sqrt{4A_2 \alpha^2 (1+\beta^2) + 1}} + \frac{2B_2(1-B_2)}{\sqrt{(2A_2+2C_2)\alpha^2 (1+\beta^2) + 1}} +$$

$$\frac{(1-B_2)^2}{\sqrt{4C_2 \alpha^2 (1+\beta^2) + 1}} + \frac{B_2^2}{\sqrt{4A_2 \alpha^2 + 1}} + \frac{2B_2(1-B_2)}{\sqrt{(2A_2+2C_2)\alpha^2 + 1}} +$$

$$\frac{(1-B_2)^2}{\sqrt{4C_2 \alpha^2 + 1}} - \frac{4}{\pi} \left(\frac{B_2^2 \arccot(B\sqrt{T_A})}{\sqrt{T_A}} + \frac{(1-B_2)^2 \arccot(\beta\sqrt{T_C})}{\sqrt{T_C}} + \frac{B_2(1-B_2) \arccot(\beta\sqrt{T_{AC}})}{\sqrt{T_{AC}}} + \right.$$

$$\left. \frac{B_2(1-B_2) \arccot(\beta\sqrt{T_{CA}})}{\sqrt{T_{CA}}} + \frac{4}{\pi} \arctan \beta - B_2 \left(\frac{2 - \frac{4}{\pi} \arccot(\beta\sqrt{2A_2 \alpha^2 (1+\beta^2) + 1})}{\sqrt{2A_2 \alpha^2 (1+\beta^2) + 1}} + \right. \right.$$

$$\left. \frac{2 - \frac{4}{\pi} \arccot(\beta\sqrt{2A_2 \alpha^2 + 1})}{\sqrt{2A_2 \alpha^2 + 1}} \right) - (1-B_2) \left(\frac{2 - \frac{4}{\pi} \arccot(\beta\sqrt{2C_2 \alpha^2 (1+\beta^2) + 1})}{\sqrt{2C_2 \alpha^2 (1+\beta^2) + 1}} + \right.$$

$$\left. \frac{2 - \frac{4}{\pi} \arccot(\beta\sqrt{2C_2 \alpha^2 + 1})}{\sqrt{2C_2 \alpha^2 + 1}} \right) \quad (19)$$

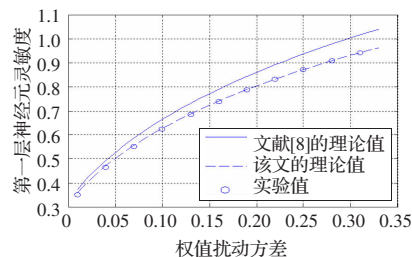


图2 第一层神经元灵敏度与权值扰动方差关系图

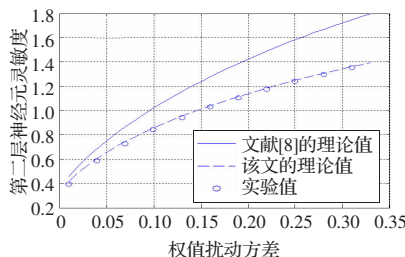


图3 第二层神经元灵敏度与权值扰动方差关系图

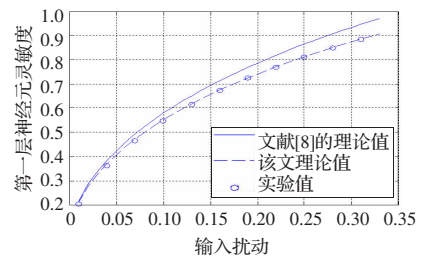


图4 第一层神经元灵敏度与输入扰动方差关系图

其中:

$$T_A = 4\beta^2 A_2^2 \alpha^4 + 2A_2 \alpha^2 + 2A_2 \alpha^2 (1+\beta^2) + 1$$

$$T_C = 4\beta^2 C_2^2 \alpha^4 + 2C_2 \alpha^2 + 2C_2 \alpha^2 (1+\beta^2) + 1$$

$$T_{CA} = 4\beta^2 C_2 A_2 \alpha^4 + 2C_2 \alpha^2 + 2A_2 \alpha^2 (1+\beta^2) + 1$$

$$T_{AC} = 4\beta^2 A_2 C_2 \alpha^4 + 2A_2 \alpha^2 + 2C_2 \alpha^2 (1+\beta^2) + 1$$

由定义1,可得第 l 层第 i 个神经元的灵敏度为:

$$S = \sqrt{\frac{D(\Delta y_j^l)}{D(y_j^l)}} \quad (20)$$

式(20)中输出方差和输出扰动方差分别由式(16),式(19)获得。

这样就求得了使用新逼近函数时,单个神经元的灵敏度公式。整个计算过程无需考虑输入和权值扰动的大小,故取消了文献[3]中的限制。应该指出,对于阈值型的激活函数灵敏度公式仍是可用的,将公式中 A_1, B_1, C_1, A_2, C_2 趋至正无穷大即可解得对应于阈值型激活的灵敏度解析式。

对于整个前馈网络的灵敏度,因为前一层的输出和输出误差是作为下一层的输入和输入误差的,所以可以采用迭代方法从输入层开始经层层前进,最终获得输出层的各神经元灵敏度即整个网络的灵敏度。

5 实验与对比

通过实验对比,考察新的灵敏度计算公式的正确性与精确性。

实验1 以常见的三层前馈神经网络为例,各层神经元个数均为40,激活函数的倾斜因子为0.5。网络输入看作是[-1,1]上均匀分布的随机变量,权值看作是[-1,1]上均匀变化的随机变量,考虑较大扰动情况,假定输入扰动为[-0.3,0.3]上均匀变化的随机变量,权值扰动作为待变量。仿真实验中,网络的输入和权值以及它们的扰动均按要求随机产生,每一个实验点的结果由10000次随机模拟实验产生。图2~3是神经元灵敏度的理论值和实验值与权值扰动方差之间的关系。

实验2 以三层前馈神经网络为例,各层神经元个数均为30,激活函数的倾斜因子为0.6。网络输入和权值的假设同实验1,输入扰动为待变量,权值扰动为[-0.1,0.1]上均匀变化的随机变量。每一个实验点的结果由10000次随机模拟实验产生。图4~5是神经元灵敏度的理论值和实验值与输入扰动方差之间的关系。

由实验1和实验2可见,随着输入扰动方差和权值扰动方差的增大,第一层和第二层神经元灵敏度也逐渐增加;计算公式的理论值与实验值十分吻合,而文献[8]出现了较大偏差,尤其是当扰动方差较大时更显示出该文理论值的准确性。

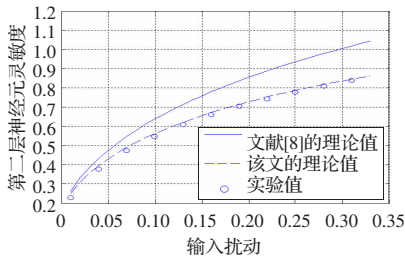


图5 第二层神经元灵敏度与输入扰动方差关系图

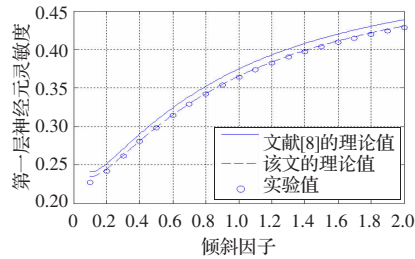


图6 第一层神经元灵敏度与倾斜因子关系图

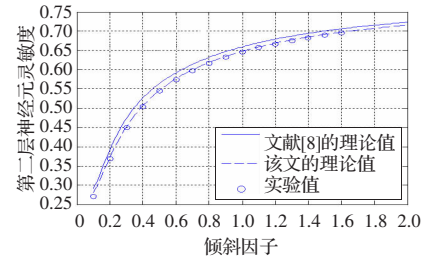


图7 第二层神经元灵敏度与倾斜因子关系图

实验3 以三层前馈神经网络为例,各层神经元个数均为40,激活函数的倾斜因子为待变量。网络输入看作是 $[-1, 1]$ 上均匀分布的随机变量,权值是 $[-5, 5]$ 上均匀变化的随机变量,输入扰动为 $[-0.2, 0.2]$ 上均匀变化的随机变量,权值扰动为 $[-0.5, 0.5]$ 上均匀变化的随机变量。图6~7是神经元灵敏度的理论值和实验值与激活函数的倾斜因子之间的关系。

可见,文献[8]的理论值在实验中与实验值的偏差不大,但该理论值与实验值吻合度仍更好。综合实验1~3,充分体现了该文理论值的准确性与精确性,这正是使用高精度逼近函数带来的效果。

6 结论

基于Piché的随机模型,通过使用两个新的逼近函数,获得了更加精确的神经网络灵敏度计算公式,尤其在输入扰动和权值扰动较大时体现了公式的高精度,克服了其他方法出现计算误差较大的缺点,对神经网络灵敏度的精密计算和应用具有一定的实际意义。

参考文献:

- [1] Zurada J M, Malinowski A, Usui S. Perturbation method for deleting redundant inputs of perceptron networks[J]. *Neurocomputing*, 1997, 14: 177-193.
- [2] Engelbrecht A P. A new pruning heuristic based on variance analysis of sensitivity information[J]. *IEEE Trans on Neural Networks*,

2001, 12(6):1386-1399.

- [3] Piché S W. The selection of weight accuracies for Madalines[J]. *IEEE Trans Neural Networks*, 1995, 6(2):432-445.
- [4] Wing W Y N. Selection of weight quantisation accuracy for radial basis function neural network using stochastic sensitivity measure[J]. *IEE Electronic Letters*, 2003, 39(10):787-789.
- [5] Engelbrecht A P, Cloete I. Incremental learning using sensitivity analysis[J]. *IEEE Trans on Neural Networks*, 1999, 2:1350-1355.
- [6] Wing W Y N, Yeung D S, Cloete I. Quantitative study on effect on center selection to RBFNN classification performance[C]//*IEEE International Conference on Systems, Man and Cybernetics*, 2004, 4: 3692-3697.
- [7] Wing W Y. A study of the difference between partial derivative and stochastic neural network sensitivity analysis for applications in supervised pattern classification problems[C]//*Machine Learning and Cybernetics*, 2004:4283-4288.
- [8] Yeung D S, Sun Xue-quan. Using function approximation to analyze the sensitivity of MLP with antisymmetric squashing activation function[J]. *IEEE Trans NN*, 2002, 13(1):34-44.
- [9] Zeng Xiao-qin, Yeung D S. Sensitivity analysis of multilayer perceptron to input and weight perturbations[J]. *IEEE Trans on Neural Networks*, 2001, 12(6):1358-1366.
- [10] Ng W W Y. Statistical output sensitivity to input and weight perturbations of radial basis function neural networks[C]//*IEEE Proc Int Conf of SMC*, 2002:503-508.

(上接12页)

从而可知,按EDD排程时的总成本最低为171960元,其完成日期分别为5、16、14、10、19、21;SPT时的总成本最高为260660元,比最低时增加约50%,其完成日期分别为5、9、13、18、21、16。OP、EDD和ESD满足所有交货期要求,OP时总成本为171930,其完成日期分别为10、23、19、14、23、25。

6 结论

提出了基于订单优先准则的装配模拟排产方法,建立了在多装配线下的订单装配排产成本模型,依据订单的重要与否分别采用后向与前向排产方法来确定订单的交货期,对电子装配企业的订单排产具有一定的指导意义。实例表明提出方法的有效性,能为企业带来成本的很大节约,保证了重要订单的准时交货,对企业来讲具有重要的战略意义。

参考文献:

- [1] Leung J Y T, Li H, Pinedo M. Scheduling orders for multiple pro-

duct types with due date related objectives[J]. *European Journal of Operational Research*, 2006, 168(2):370-389.

- [2] Leung J Y T, Li H, Pinedo M. Scheduling orders for multiple product types to minimize total weighted completion time[J]. *Discrete Applied Mathematics*, 2007, 155(8):945-970.
- [3] Li K, Ganesan V K, Sivakumar A I. Scheduling of single stage assembly with air transportation in a consumer electronic supply chain[J]. *Computers & Industrial Engineering*, 2006, 51(2):264-278.
- [4] Lin B M T, Kononov A V. Customer order scheduling to minimize the number of late jobs[J]. *European Journal of Operational Research*, 2007, 183(2):944-948.
- [5] Wang G, Cheng T C E. Customer order scheduling to minimize total weighted completion time[J]. *Omega-International Journal of Management Science*, 2007, 35(5):623-626.
- [6] 徐俊刚,戴国忠,王宏安.生产调度理论和方法研究综述[J]. *计算机研究与发展*, 2004, 41(2):257-267.