最优加权系数的神经优化方法

宋玉阶　　吴怀宇

（武汉冶金科技大学自动化系　武汉 430081）

摘　要：提出了利用神经网络求解最小二乘准则下和最小一乘准则下最优加权系数的方法，这种神经优化方法可以在电路时间常数数量级给出准确解，而且不存在编程复杂性问题。

关键词：最小二乘准则　最小一乘准则　最优加权系数　神经网络

0　引　　言

客观实际中的非线性系统，往往内部结构复杂、输入输出变量众多，采用单个的模型或部分的因素和指标仅能体现系统的局部，多个模型的有效组合或多个变量的科学综合能显著地提高预测精度与模拟评价效果。

任何数学模型都是对实际系统的某种简化和抽象，描述同个系统的多种模型往往各有特点和利弊，合理综合可获博采众长的效果。

权重综合和区域综合即是比较常用的两种方法，权重综合中除算术平均法、标准差法、方差倒数法、均方倒数法、离异系数法、二项式系数法、简单加权法、三点法、层次分析法、德尔菲法外，最常用的方法为最优加权法[1]，该法的原理是依据某种最优准则构造目标函数 Q，在一定约束条件下通过极小化 Q 以求得权系数，但该优化问题的数学计算中含有大量求逆运算。

由于近来对神经网络深入和广泛的研究，人们提出了几种利用神经网络求解优化问题的方法，Hopfield 于 1982 年用 HNN（Hopfield Neural Network）模型[2]，成功地解决了计算复杂度为 NPC 的旅行商问题，并用硬件实现了该模型。文献[3]基于 Lagrange 乘子理论通过构造原问题的 Lagrange 函数，提出 LPNN（Lagrange Programming Neural Network）网络模型，比较成功地解决了约束优化问题。

作者根据 ANN（Artificial Neural Network）的基本优化机理，研究了基于 Lagrange 函数适合求解最优加权系数的神经优化方法。

1　最优加权模型

设 $\{x_i\}$，$i=1,2,\cdots,N$ 为观测序列，共有 J 个数学模型对之拟合描述，拟合误差 e_i，$i=1,2,\cdots,N$ 为

$$e_i = x_i - \hat{x}_i = x_i - \sum_{j=1} w_j\hat{x}_i(j) = \sum_{j=1}^{J}(x_i - \hat{x}_i(j)) = \sum_{j=1}^{J} w_j e_i(j) \tag{1}$$

式中 $\hat{x}_i(j)$、$e_i(j)$ 为第 J 个模型对 x_i 的拟合值和拟合误差。

若取 e_i 为误差统计量，此时最小二乘准则下综合模型权系数和最小一乘准则下综合模型权系数的最优解分别由以下模型(2)、(3)获得

$$\begin{cases} \min & Q = \sum_{i=1}^{N} e_i^2 \\ \text{s.t} & \sum_{j=1}^{n} w_j = 1,\ w_j \ge 0,\ j = 1,2\cdots,J \end{cases} \tag{2}$$

$$\begin{cases} \min & Q = \sum_{i=1}^{N} |e_i| \\ \text{s.t} & \sum_{j=1}^{J} w_j = 1,\ w_j \ge 0,\ j = 1,2\cdots,J \end{cases}$$

式(2)中如果去掉非负权重约束条件 $w_j \ge 0, j=1,2\cdots,J$，可利用 Lagrange 乘子法求得解析解，最优权向量 $W_0 = (R^TE^{-1}R)^{-1}E^{-1}R$，最小 Q 值 $Q_0 = (R^TE^{-1}R)^{-1}$。由于该法处理优化问题式(2)时存在大量求逆运算，而且实际问题往往不允许出现负权重，因此，式(2)的优化问题的数学计算只能利用数值方法迭代处理。

式(3)只能利用迭代方法进行计算,由于式(3)计算的复杂性,更多场合人们乐意采用最小二乘准则下式(2)计算最优加权系数,但一般情况下,最小一乘比最小二乘的结果稳健。利用神经网络可以对式(2)和式(3)之类的优化问题进行优化计算。

2 神经优化计算

神经网络是由神经元组成的一个非线性大规模动力学系统。根据动力学系统的观点,系统的最终行为完全由它的吸引子决定。对于稳定系统而言,给于系统一定的输入,系统在自身动力学系统的作用下,最终会流向其稳定的吸引子。如果将网络系统运动的最终稳定吸引子考虑为实际问题的最优解,那么,神经网络由某个初始状态出发而达到稳定状态的演化过程,就是在可行域内搜索目标函数最优值的过程。

2.1 最小二乘准则下综合模型权系数的神经优化计算

为求解最小二乘准则下优化问题式(2),需将式(2)表示为矩阵形式,令 $R=(1,1,\cdots,1)^T$, $W=(W_1,W_2,\cdots,W_J)^T$, $e=(e_1,e_2,\cdots,e_N)^T$, $e_j=(e_1(j),e_2(j),\cdots e_2(j),\cdots e_N(j))^T$, $j=1,2\cdots,J$, $E=(e_{ij})_{J\times J}$, $e_{ij}=e_i^Te_j=\sum_{k=1}^{N}e_k(i)e_k(j)$, E 称为信息阵,为对称正定矩阵,由此可得

$$e_{ij}=x_i-\hat{x}i=\sum_{j=1}^{J}w_ie_i(j)=(e_i(1),e_i(2),\cdots e_i(J))W$$

$$e=\begin{Bmatrix}e_1\\ \vdots\\ e_N\end{Bmatrix}=\begin{Bmatrix}e_1(1)\cdots e_1(J)\\ \vdots\ \vdots\ \ \vdots\\ e_N(1)-e_N(J)\end{Bmatrix}W=(e_1,e_2,\cdots e_J)W$$

则式(2)的矩阵形式为

$$\begin{cases}\min & Q=e^Te=W^TEW\\ s.t & R^TW=1,w_j\geq 0,i=1,2\cdots,J\end{cases}\tag{4}$$

为了求解最小二乘准则下优化问题式(4),得到最优加权系数建立动力学系统

$$\begin{cases}\dfrac{dU}{dt}=-[2EW-K_1R_y+k_2R(R^TW-1)]\\ \dfrac{dv}{dt}=-(R^TW-1)\\ w_j=g(u_j), \qquad j=1,2,\cdots,J\\ y=g(v)\end{cases}\tag{5}$$

其中,$W\in R^J$ 是网络的输出,对应优化变量。相应于(5)式的神经网络的电路框图如文献[4]图1所示。

由图可见,神经网络的电路结构比较简单,只需构造 J+1 个神经单元,连接权益共有个 J^2+2J+1, $g(\cdot)$是神经元激活函数,取 Sigmoid 函数。由于 Sigmoid 函数为非负单增函数,可自然满足约束条件 $W_J\geq 0, J=1,2,\cdots,J$。

神经网络(5)式是全局稳定的,并且收敛到问题(4)式的最优解。

选取动力学系统(5)式的能量函数为

$$E(w,y)=W^TEW-k_1y(R^TW-1)+\frac{k_2}{2}\|R^TW-1\|^2\tag{6}$$

则能量函数 E 对时间的变化率为

$$\frac{dE}{dt}=\sum_{j=1}^{J}\frac{\partial E}{\partial w_j}\cdot\frac{dw_j}{dt}+\frac{\partial E}{\partial y}\cdot\frac{dy}{dt}=\sum_{j=1}^{J}\frac{\partial E}{\partial w_j}\cdot\frac{\partial w_j}{\partial u}\cdot\frac{du}{\partial t}+\frac{\partial E}{\partial y}\cdot\frac{\partial y}{\partial v}\cdot\frac{dv}{dt}=[\nabla_wE(w,y)]^T\cdot G_u\cdot\frac{dU}{dt}+g(v)[\nabla_yE(x,y)]^2\leq 0\tag{7}$$

$$g'[\nabla_wE(w,y)]^T\cdot\nabla_wE(w,y)-g'(v)[\nabla_yE(x,y)]^2\leqslant 0$$

偏导数,且有$\frac{dU}{dt}=-\nabla_wE(w,y)$, $\frac{dv}{dt}=\cdot-\nabla_yE(w,y)$, $G_u=\mathrm{diag}(g(u_1),g(u_2),\cdots,g(u_J))$, $g(u_j)>0$, $g(v)>0$, $j=1,2,\cdots,J$。

因此,系统的能量函数 E(w,y)随时间的增加而不断减小,直到系统的稳定状态为止。所以(6)式是该动力学系统的 Lyapunov 函数,且动力学系统是全局稳定的。

根据(7)式,当且仅当 dU/dt=0, dv/dt=0,才有 dE/dt=0,说明该动力学系统进入稳定状态时,能量函数 E

(w,y)不再变化(不再减小),$E(w,y)$取得局部极小值。

由于最小二乘准则下综合模型权系数的最优解优化问题是一凸性优化问题,信息阵 E(实为协方差矩阵)为对称正定矩阵,$Q=W^TEW$ 在开凸集 $W\in R^J$ 上是一个一次可微函数,Hessian 矩阵 $H(w)=E$ 对于每一个 $w\in W$ 是正定的,$Q=W^TEW$ 在 W 上是一个严格凸函数,因此,问题(4)式的每一局部解 W^* 即为全局最优解,且全局最优解的集合 S 为凸集。

另外,根据 Kuhn—Tucker 最优定理,$dU/dt=0$,$dv/dt=0$,满足 $W=W^*$ 是(4)式全局最优解的充要条件。因此,动力学系统的稳定平衡点 $dU/dt=0$,$dv/dt=0$,对应于问题(4)式的唯一全局最优解,$W_j^*=g(u_j^*)$,$j=1,2,\cdots J$。不管系统初始状态如何,系统都将演化到平衡态。亦即不管优化变量初始值为何,神经网络都将使变量收敛到全局唯一最优解。

2.2 最小一乘准则下综合模型权系数的神经优化计算

为了利用神经网络求解式(3),先将(3)式进行数学处理。因为 $e=(e_1,e_2,\cdots,e_N)$,$e_i=\sum_{j=1}^{J}w_je_i(j)$,$Q=\sum_{i=1}^{N}|e_i|$,所以,$Q=\|e\|_1$。目前处理 l_1 范数形式的优化问题的神经网络既少且不成熟,必须将该优化问题转化成其它形式优化问题。根据许瓦尔兹(schwartz)不等式,

$Q=\sum_{i=1}^{N}\left|\sum_{j=1}^{J}w_je_i(j)\right|\le\sum_{i=1}^{N}\sum_{j=1}^{J}|w_je_i(j)|=\sum_{i=1}^{N}\sum_{j=1}^{J}w_j|e_i(j)|=\sum_{j=1}^{J}w_j|(e_i(j)|=\sum_{j=1}^{J}w_j\sum_{i=1}^{N}|e_i(j)|=\sum_{j=1}^{J}s_jw_j=S^TW$ 其中,$s_j=\sum_{j=1}^{N}|e_i(j),|$,$S=(S_1,S_2,\cdots,S_J)^T$

等效优化数学模型为(8)式,式(8)为线性优化问题,其最优解是式(3)的近似最优解。

$$\begin{cases}\min & Q=S^TW\\ S.t & R^TW=1,w_j\ge0,j=1,2\cdots,J\end{cases}\tag{8}$$

为了求解优化问题(8)式,根据罚函数收敛定理[5],构造动力学系统状态方程为

$$\begin{cases}\dfrac{dU}{dt}=-[S+k_1W+k_2R(R^TW-1]\\ w_j=g(u_j),\quad j=1,2,\cdots,J\end{cases}\tag{9}$$

相应于稳定的动力学系统式(9)的能量函数为

$$E(w)=SW+\frac{k_1}{2}\|W\|^2+\frac{k_2}{2}\|R^TW-1\|^2\tag{10}$$

3 模拟计算实例

储蓄存款余额最优加权组合预测的最优加权系数计算[1],用最小平方法和三次指数平滑法对 1971~1982 年乡村居民标准储蓄存款余额进行预测,两种模型的预测值与实际值列于下表:

	1	2	3	4	5	6	7	8	9	10	11	12
x_i	11.49	13.06	15.34	20.58	23.28	26.46	27.33	34.22	40.19	53.37	77.7	100.63
$x_i^{(1)}$	18.47	14.54	12.84	13.38	16.15	21.16	28.40	37.87	49.58	63.53	79.00	98.12
$x_i^{(2)}$	10.03	11.23	15.24	18.67	27.78	26.36	29.67	27.40	42.73	47.36	71.00	109.32

3.1 最小二乘准则下最优加权系数模拟计算

根据式(5),信息阵 $E=\begin{bmatrix}401.5570 & -114.6634\\ -114.6634 & 245.5786\end{bmatrix}$,取 $k_1=500$,$k_2=200$,神经元激活函数为 sigmoid 函数,则

状态方程为 $\begin{cases}du_1/dt=-[1003.114w_1-29.3268w_2+200+500y]\\ du_2/dt=-[691.157w_2-29.3268w_1+200+500y]\\ dv/dt=-500(w_1+w_2-1)\end{cases}$

(下转第 25 页)