

文章编号:1673-5005(2009)06-0095-05

一种基于综合目标函数的神经网络学习算法

徐宝昌¹, 罗雄麟¹, 王金山²

(1. 中国石油大学机电工程学院, 北京 102249; 2. 塔里木油田公司, 新疆 库尔勒 841000)

摘要:为提高多层前向神经网络的学习速度和算法的稳定性,提出一种基于综合目标函数的改进学习算法。该算法在误差平方和目标函数中引入一个辅助约束项构成综合目标函数,并利用综合目标函数训练网络的输出层权值,采用牛顿法推导出训练输出层权值的递推公式。辅助约束项隐含有对网络输出平滑性的约束,提高了学习算法的稳定性。利用该算法对不同非线性函数生成的样本数据的学习结果表明,新算法的收敛速度、精度均优于 Karayiannis 等人的二阶学习算法。

关键词:神经网络; 学习算法; 综合目标函数

中图分类号:TP 183 **文献标识码:**A

A novel neural network training algorithm based on generalized objective function

XU Bao-chang¹, LUO Xiong-lin¹, WANG Jin-shan²

(1. Faculty of Mechanical and Electronic Engineering in China University of Petroleum, Beijing 102249, China;

2. PetroChina Tarim Oilfield Company, Korla 841000, China)

Abstract: A novel training algorithm was proposed to improve the learning rate and stability of the multi-layer feedforward neural networks. The generalized objective function was constructed by adding an auxiliary constraint term to the sum of the squared errors in the algorithm. The weight matrix of output layer was trained using the generalized objective function. The recursive equations for training the weight matrix of output layer were derived using Newton iterative algorithm without any simplification. The auxiliary constraint term involves the requirement for the smoothness of output which could improve the stability of the algorithm. The high-order derivative information of the neuron action function was used during the training procedure, so the algorithm had high convergence speed. In the end, the algorithm was used to learn training pattern of different nonlinear function. Simulation results show that the convergent rate and accuracy of the algorithm are better than those of the Karayiannis's second-order learning algorithm.

Key words: neural network; training algorithm; generalized objective function

多层前向神经网络是应用最普遍的一种神经网络,在模式识别、信号处理、函数逼近、系统辨识和优化等方面得到广泛的应用^[1,2]。Yam 等利用最小二乘法来确定前馈网络的初始权值^[3],谢富强等对现有多种前向网络的权值初始化方法进行了综述,提出了若干待研究的问题^[4]。更多的研究集中在训练网络权值的优化方法上,如基于最速下降法的经典 BP 算法^[7],基于模拟退火算法和遗传算法的全

局随机学习算法^[5],基于 BFGS 优化方法的学习算法^[6],基于最优学习速率的 BP 学习算法^[7]。除了上述因素外,学习算法的性能还受到目标函数形式的影响,Jose 等采用均方误差(mean square)作为目标函数,并利用遗传算法来进行网络权值的训练^[8],Billing 等采用预报误差作为目标函数并推导出多层前向网络的二阶学习算法,其收敛速度优于 BP 算法^[9]。目前已有的神经网络的缺点在于单一

收稿日期:2009-04-12

基金项目:中石油重点科技开发项目(2008C-2203)

作者简介:徐宝昌(1974-),男(汉族),辽宁凌源人,副教授,博士,主要从事复杂油气系统的信息处理、建模与先进控制研究工作。

的目标函数并不能够保证在整个学习过程均保持较快的收敛速度,例如二阶学习算法在目标函数二次性较强的区域有较快收敛速度,而在其他区域收敛较慢,算法稳定性不好^[10]。笔者提出一种新型的综合目标函数,利用综合目标函数来训练网络的输出层权值,并利用牛顿法推导基于综合目标函数的输出层权值计算的递推公式,改善算法对样本噪声的鲁棒性,提高算法的学习速度和学习精度。

1 新型神经网络学习算法

神经网络的实际输入输出样本数据在经过一定的数据预处理后,仍然会受到一些随机噪声的影响,当采用误差平方和这样的二次误差函数进行训练时,这些噪声会使网络在训练过程中出现病态数据(矩阵奇异),从而影响算法的稳定性和收敛性。针对这一问题,POGGIO等在径向基函数神经网络的训练过程中引入了正规化理论,来求解径向基函数神经网络的函数逼近问题,并得到了稳定解^[11]。笔者在多层前向神经网络的训练过程中,借鉴正规化理论的思想,在误差平方和目标函数中引入一个辅助约束项构成综合目标函数,引入的辅助项隐含有对网络输入输出数据平滑性的约束,一定程度上利用了输入输出数据的先验性信息,能够稳定网络训练过程中权值的递推计算,提高算法的学习速度和学习精度。

考虑具有 r_n 个隐层的多层前向神经网络。训练样本为 (X_k^r, Y_k) , $k = 1, \dots, m$; 输入向量 $X_k^r = (x_{1k}, x_{2k}, \dots, x_{n_k})^T$, 期望输出向量 $Y_k = (y_{1k}, \dots, y_{n_{ok}})^T$, 对 $\forall k$, 神经网络输出层和各隐层输出向量分别为 $\hat{Y}_k = (\hat{y}_{1k}, \dots, \hat{y}_{n_{ok}})^T$, $X_k^{r-1} = (\hat{h}_{1k}^r, \hat{h}_{2k}^r, \dots, \hat{h}_{n_k}^r)^T$, $r = 1, 2, \dots, r_n$ 。

式(1)中神经网络的输入输出关系为

$$\hat{y}_{ik} = \sigma_0(\bar{y}_{ik}) = \sigma_0((W_i^0)^T X_k^0) = \sigma_0\left[\sum_{j=0}^{n_1} w_{ij}^0 \hat{h}_{jk}^1\right],$$

$$\hat{h}_{jk}^r = \sigma_r(\bar{h}_{jk}^r) = \sigma_r((W_j^r)^T X_k^r) = \sigma_r\left[\sum_{v=0}^{n_{r+1}} w_{jv}^r \hat{h}_{vk}^{r+1}\right].$$

其中当 $r = r_n$ 时, $\hat{h}_{vk}^{r+1} = x_{vk}$, $v = 1, 2, \dots, n_i, n_{r+1} = n_i$, n_i 为输入层神经元个数, $\sigma_r(\cdot)$ 采用双曲正切函数。

引入辅助约束项的综合目标函数为

$$E_m = \frac{1}{2} \sum_{i=1}^{n_o} \sum_{k=1}^m e_{ik}^2 + \frac{1}{2} \lambda \sum_{i=1}^{n_o} \sum_{k=1}^m (\hat{y}'_{ik})^2. \tag{3}$$

式中, n_o 为输出神经元个数; 第一项为误差平方和函数, 第二项为非负的辅助约束项; \hat{y}'_{ik} 为网络输出对输出层权值的一阶导数; $\lambda > 0$, 为平滑性参数, 反映辅助约束项在综合目标函数中所占的比重。

式(3)中的第二项反映了对网络输出平滑性的约束, 而这个约束是由网络输出对输出层权值的一阶导数来进行量化和描述的, 即对网络输出平滑性的约束可以通过在线调整输出层的权值来实现。因此, 本文算法首先以式(3)为目标函数, 利用最优化方法来得到输出层权值的递推计算公式。为了后面推导的方便, 将式(3)化简为如下形式:

$$E_m = \sum_{i=1}^{n_o} E_{i,m} = \sum_{i=1}^{n_o} \left[E_{i,m-1} + \frac{1}{2} e_{im}^2 + \frac{1}{2} (\hat{y}'_{im})^2 \right]. \tag{4}$$

其中

$$E_{i,m-1} = \frac{1}{2} \sum_{k=1}^{m-1} e_{ik}^2 + \frac{1}{2} \lambda \sum_{k=1}^{m-1} (\hat{y}'_{ik})^2.$$

从式(4)可以看出, $E_{i,m-1}$ 只与样本 (X_k^r, Y_k) , $k = 1, 2, \dots, m-1$ 有关。本文算法推导的目的就是利用新样本 (X_m^r, Y_m) , 通过最优化方法迭代计算使式(4)最小化的网络权值, 以得到递推形式的网络学习算法。

应用牛顿法求解式(4)的极小化问题, 得到输出层权值的递推计算形式为

$$W_{i,m}^0 = W_{i,m-1}^0 + \Delta W_{i,m}^0, \tag{5}$$

其中

$$\Delta W_{i,m}^0 = -\alpha (H_{i,m}^0)^{-1} G_{i,m}^0.$$

式中, $W_{i,m}^0$ 表示输出层第 i 个神经元所对应的权值向量; α 为学习速率; $G_{i,m}^0$ 和 $H_{i,m}^0$ 分别为 $E_{i,m}$ 关于 $W_{i,m}^0$ 的梯度和 Hessian 阵的值。

利用矩阵微分可得

$$G_{i,m}^0 = -\sum_{k=1}^m \sigma'(\bar{y}_{ik}) e_{ik} X_k^0 + \lambda \sum_{k=1}^m (X_k^0)^T X_k^0 \sigma'(\bar{y}_{ik}) \times \sigma''(\bar{y}_{ik}) X_k^0 = G_{i,m-1}^0 + \varepsilon_{im} X_m^0 + \lambda \gamma_{im} X_m^0. \tag{6}$$

其中

$$\varepsilon_{im} = -\sigma'(\bar{y}_{im}) e_{im},$$

$$\gamma_{im} = (X_m^0)^T X_m^0 \sigma'(\bar{y}_{im}) \sigma''(\bar{y}_{im}).$$

以式(6)为基础, 可以进一步得到 Hessian 阵的计算结果

$$H_{i,m}^0 = H_{i,m-1}^0 + z_{im} X_m^0 (X_m^0)^T, \tag{7}$$

其中

$$z_{im} = [(\sigma'(\bar{y}_{im}))^2 - \sigma''(\bar{y}_{im}) e_{im}] + \lambda [(\sigma''(\bar{y}_{im}))^2 + \sigma'(\bar{y}_{im}) \sigma'''(\bar{y}_{im})] (X_m^0)^T X_m^0.$$

令 $P_{im}^0 = [H_{i,m}^0]^{-1}$, 由式(7)和矩阵求逆理可

得^[12]

$$\begin{aligned} P_{im}^0 &= [(P_{i,m-1}^0)^{-1} + z_{im} X_m^0 (X_m^0)^T]^{-1} = \\ P_{i,m-1}^0 - M_{im} (X_m^0)^T P_{i,m-1}^0 &= [I - M_{im} (X_m^0)^T] P_{i,m-1}^0, \end{aligned} \quad (8)$$

其中

$$\begin{aligned} M_{im} &= P_{i,m-1}^0 X_m^0 [\beta_{im} + (X_m^0)^T P_{i,m-1}^0 X_m^0]^{-1}, \\ \beta_{im} &= 1/z_{im}. \end{aligned}$$

将 M_{im} 两边右乘 $[\beta_{im} + (X_m^0)^T P_{i,m-1}^0 X_m^0]$, 整理得

$$\beta_{im} M_{im} = P_{i,m-1}^0 X_m^0 - M_{im} (X_m^0)^T P_{i,m-1}^0 X_m^0. \quad (9)$$

将式(8)两边右乘 X_m^0 , 整理得

$$P_{i,m}^0 X_m^0 = P_{i,m-1}^0 X_m^0 - M_{im} (X_m^0)^T P_{i,m-1}^0 X_m^0. \quad (10)$$

比较式(9),(10)可得

$$P_{i,m}^0 X_m^0 = \beta_{im} M_{im}. \quad (11)$$

把式(6)和(8)代入式(5),并注意式(11)得

$$\begin{aligned} \Delta W_{ik} &= -[H_{im}^0]^{-1} G_{i,m}^0 = -P_{i,m}^0 [G_{i,m-1}^0 + \varepsilon_{im} X_m^0 + \\ &\lambda \gamma_{im} X_m^0] = -[I - M_{im} (X_m^0)^T] P_{i,m-1}^0 G_{i,m-1}^0 - \\ &P_{i,m}^0 [\varepsilon_{im} X_m^0 + \lambda \gamma_{im} X_m^0] = [I - M_{im} (X_m^0)^T] \Delta W_{i,m-1}^0 - \\ &P_{i,m}^0 [\varepsilon_{im} X_m^0 + \lambda \gamma_{im} X_m^0] = \Delta W_{i,m-1}^0 - M_{im} [\beta_{im} \varepsilon_{im} + \\ &(X_m^0)^T \Delta W_{i,m-1}^0 + \lambda \beta_{im} \gamma_{im}]. \end{aligned} \quad (12)$$

其中

$$P_{i,0}^0 = \alpha I.$$

α 为一个很大的数,这里取为 10^4 。

式(5),(8),(12)即为逐步递推计算网络输出层权值的计算公式,每输入一个样本则对权值进行一次修正,当所有样本均输入一次后视作一步迭代,如果未达到精度的要求,则进入下一步迭代,直至满足精度为止。

对网络中输入层和隐含层权值的计算,本文中仍采用经典的逐步计算的BP算法来进行训练。

2 算法性能分析

(1) 与经典的BP算法相比,本文算法在最小二乘目标函数中引入了一个辅助约束项构成综合目标函数,可以提高算法对样本噪声的鲁棒性,提高算法的稳定性和匹配速度。

(2) 平滑性算子 λ 用来衡量对输出平滑性约束的程度,从式(5),(8)(12)可以看出,当 $\lambda = 0$ 时,综合目标函数等价于误差平方和目标函数, λ 越大,则新算法对输入噪声的鲁棒性越强。 λ 值的确定与样本的平滑性等先验信息有关,对于含有较大噪声的样本相应地取较大的 λ 值,随着训练的逐步进行,按照一定规律减小 λ 值可以取得较好的训练结果。

(3) 本文算法充分利用了综合目标函数的一

阶、二阶导数信息,在推导过程中未作任何简化,实现了Newton搜索方向和Hessian阵逆的递推运算,具有较好的收敛性能,匹配速度快;算法采用递推计算形式,有助于减小计算的存储量和复杂度。

(4) 本文算法同时利用 γ_{im} , z_{im} 和 ε_{im} 来进行权值的修正,比经典BP算法用到了更多的信息,这将更有利于NN权值的调整。当神经元作用函数取双曲正切函数时,目标函数的高阶导数计算并不复杂,计算量也不大。

(5) 本文算法采用了混合优化策略的学习算法,在权值训练的过程中,当远离极值点时,BP算法的收敛速度会很快,权值训练可以着重调整输入层和隐含层的权值,BP算法会起主要作用,而输出层权值的调整则着重于稳定算法的运行,在满足网络输出平滑性要求方面起到辅助作用。当网络训练接近极小点时,由于输出层权值是采用二阶学习算法进行训练,在接近极小点附近具有较快的收敛速度,此时输出层权值的训练会占主要地位,仍能够保证较快的训练速度。通过在不同的网络层次采用不同的算法训练权值,综合利用了不同算法在不同区域的优点,能够提高整个训练过程的速度。

(6) 权值增量的递推公式还可以写成

$$\Delta W_{i,m}^0 = [I - M_{im} (X_m^0)^T] \Delta W_{i,m-1}^0 - P_{i,m}^0 X_m^0 [\varepsilon_{im} + \lambda \gamma_{im}]. \quad (13)$$

权值增量带有惯性项的BP算法公式为

$$\Delta W_{i,m} = \alpha \Delta W_{i,m-1} + \eta X_m e_{i,m}. \quad (14)$$

式中, α 为惯性因子; η 为学习速率。

比较式(13)和式(14)可知,式(13)也是一种带有惯性项的递推算法,其第一项为牛顿搜索方向的惯性项,第二项为瞬时更新项。不同之处在于式(13)实现了惯性因子和学习速率的在线自适应计算,而不用人为设定,同时式(13)中的瞬时更新项包含有 $\lambda \gamma_{im}$ 项,它反映了辅助约束项对权值训练的影响。惯性项的引入能有效防止算法产生振荡,从而加快收敛过程。

3 仿真实验结果

利用Visual Basic 6.0编制本文算法的仿真程序,并对不同的样本数据进行训练,以训练的速度和精度作为评价指标,并和Karayiannis所给出的二阶学习算法(一种基于最小二乘目标函数和近似牛顿法推导的算法,具备比经典BP算法更好的学习速度和精度)的训练结果进行比较,以验证本文算法的性能。在进行算法比较时,用训练的迭代步数来

评价算法的训练速度,用神经网络模型输出值与真实值的平均误差来评价算法的精度。

本文的训练样本数据由 Logistic 函数、CGM 函数、正弦函数生成,并在仿真数据中分别加入 5% 和 10% 的 Gauss 噪声。其中 Logistic 和 CGM 函数是用来描述一类具有“成长”规律的系统函数。由机理分析和实测数据变化趋势分析,发现油田的累积产油量曲线呈现 S 型或拉长的 S 型,即具有明显的“成长”曲线的趋势,油田累积产油量的变化规律可

以用这两种函数来进行刻画。本文中以这两种函数所生成的样本数据进行训练,可以为将来利用神经网络模拟油田产油量这一复杂非线性函数奠定基础。

选择 4 层网络,神经网络的结构为输入层、第一隐含层、第二隐含层、输出层,其神经元个数分别为 1,10,20,1。学习速率 α 的值由一维搜索来确定。得到 Logistic 函数和正弦函数的仿真结果如图 1 和表 1 所示。

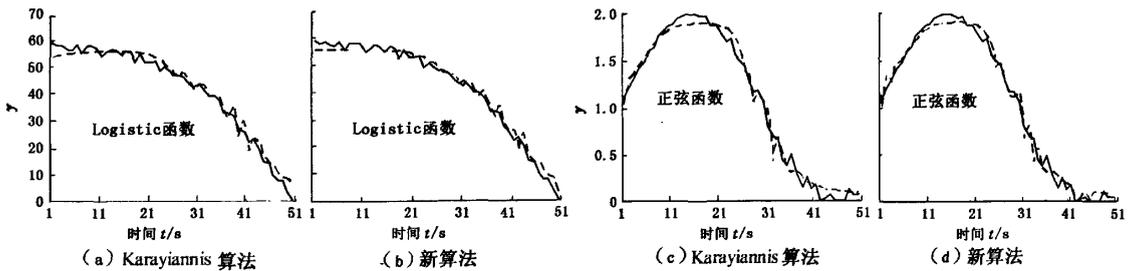


图 1 Logistic 和正弦函数加入 5% 噪声时两种算法的仿真曲线

Fig. 1 Simulation curves of two algorithms for Logistic function and sine function by adding 5% noise

表 1 加入 5% 和 10% 噪声新算法与 Karayiannis 算法的训练结果对比

Table 1 Comparison of training rate and precision between two algorithms when adding 5% and 10% noises

加入 Gauss 噪声	函数类别	迭代步数		平均误差/ 10^{-3}	
		Karayiannis 算法	新算法	Karayiannis 算法	新算法
5%	Logistic 函数	107	5	4.74	3.96
	CGM 函数	43	8	6.94	5.59
	正弦函数	512	386	4.79	3.97
10%	Logistic 函数	19	4	7.03	6.66
	CGM 函数	35	14	7.84	6.77
	正弦函数	469	387	4.99	4.99

仿真结果表明:本文算法的迭代步数要远小于 Karayiannis 算法的迭代步数,每步迭代只是增加了 β_{im} 的计算,其对每步迭代计算量的影响可忽略不计;从训练的精度看,本文算法与 Karayiannis 算法的精度相比亦有所提高。可见本文算法用于非线性模型的辨识具有较高的速度和精度。

4 结束语

给出了一种基于综合目标函数的新型神经网络学习算法,在综合目标函数中含有对样本输出的平滑性约束项,并利用牛顿法推导了算法的递推算式。提高了网络训练的稳定性、速度和精度,该学习算法无论是匹配的速度还是精度均优于已有的 Karayiannis 二阶学习算法,具有一定的实际应用前景。

参考文献:

- [1] 蔡自兴. 智能控制原理与应用[M]. 北京:清华大学出版社,2007:142.
- [2] KATHIRVALAKUMAR T, THANGAVEL P. A new learning algorithm using simultaneous perturbation with weight initialization[J]. Neural Processing Letters, 2003(17): 55-68.
- [3] YAM Y F, CHOW T W S. A new method in determining the initial weights of feedforward neural networks [J]. Neurocomputing, 1997(16):23-32.
- [4] 谢富强,唐耀庚. 多层前向神经网络权值初始化的研究进展[J]. 南开大学学报:自然科学版,2006,20(3):98-101.
XIE Fu-qiang, TANG Yao-geng. The development of weight initialization methods for MFN [J]. Journal of Nankai University (Science and Technology), 2006,20(3):98-101.
- [5] 许增福,王宏伟,吴贵生. 基于过程神经网络和量子遗传算法的油藏采收率参量逆向求解[J]. 中国石油大学学报:自然科学版,2007,31(6):120-126.
XU Zeng-fu, WANG Hong-wei, WU Gui-sheng. Converse solution of oil recovery ratio based on process neural network and quantum genetic algorithm [J]. Journal of China University of Petroleum (Edition of Natural Science), 2007,31(6):120-126.
- [6] SEAN M, GEORGE I. A variable memory quasi-Newton

- training algorithm[J]. *Neural Processing Letters*, 1999(9):77-89.
- [7] DANILO P Mandic, JONATHON A Chambers. Towards the optimal learning rate for backpropagation[J]. *Neural Processing Letters*, 2000(11):1-5.
- [8] JOSE L Sanz-gonzalez, DIEGO A, JUAN S. Importance sampling and mean-square error in neural detector training[J]. *Neural Processing Letters*, 2002(16):259-276.
- [9] BILLING S A, et al. A comparison of the back-propagation and recursive prediction error algorithm for training neural networks[J]. *Mechanical System and Signal Processing*, 1991:233-255.
- [10] 王永骥,涂健. 神经网络控制[M]. 北京:机械工业出版社,1999:56.
- [11] POGGIO T, GIROSI F. Networks for approximation and learning[J]. *Proceedings of the IEEE*, 1990(78):1481-1497.
- [12] 刘铁男,陈广义,任伟建. 时变结构系统的辨识预报和控制[M]. 哈尔滨:黑龙江科学技术出版社,1998:7.

(编辑 沈玉英)

(上接第90页)

- [6] 王宏伟,王建军,金有海. 导叶式旋风管入口环形空间内气相流场数值模拟[J]. *石油化工设备*, 2006, 35(2):33-36.
- WANG Hong-wei, WANG Jian-jun, JIN You-hai. Numerical simulation of the gas phase flow field in inlet annular space of a guide vane cyclone tube[J]. *Petro-Chemical Equipment*, 2006, 35(2):33-36.
- [7] 王建军,王连泽,刘成文. 旋风分离器排气管内流动分析及减阻机理[J]. *过程工程学报*, 2005, 5(3):251-254.
- WANG Jian-jun, WANG Lian-ze, LIU Cheng-wen. Research on the flow field in the exit tube and the Mechanism of pressure drop reduction in a cyclone separator[J]. *The Chinese Journal of Process Engineering*, 2005, 5(3):251-254.
- [8] Arman Raoufi, Mehrzad Shams, Meisam Farzaneh. Numerical simulation and optimization of fluid flow in cyclone vortex finder[J]. *Chemical Engineering Processing*, 2008, 47:128-137.
- [9] LUCÍA FERNÁNDEZ MARTÍNEZ, ANTONIO GUTIÉRREZ LAVÍN. Vortex finder optimum length in hydrocyclone separation[J]. *Chemical Engineering and Processing*, 2008, 47:192-199.

(编辑 沈玉英)

(上接第94页)

(2) 管道横截面上的速度分布是不均匀的,越接近管壁速度越小,越接近管道中心速度越大。管道横截面上的温度分布也是不均匀的,接近管壁处由于受到管壁的加热作用,温度具有较高值,接近管道中心处温度具有较低值。

参考文献:

- [1] 李治龙,钱武鼎. 我国油田用泡沫流体综述[J]. *石油钻采工艺*, 1994, 11(1):1-5.
- LI Zhi-long, QIAN Wu-ding. Survey on the application of foam fluid in China oilfield[J]. *Oil Drilling & Production Technology*, 1994, 11(1):1-5.
- [2] 李兆敏,蔡国琰. 非牛顿流体力学[M]. 东营:石油大学出版社,1998.
- [3] 廖广志,李立众. 常规泡沫驱油技术[M]. 北京:石油工业出版社,1999:65-92.
- [4] 李兆敏,孙茂盛,林日亿,等. 泡沫封堵及选择性分流实验研究[J]. *石油学报*, 2007, 28(4):115-118.
- LI Zhao-min, SUN Mao-sheng, LIN Ri-yi, et al. Laboratory study on foam plugging and selective divided-flow[J]. *Acta Petrolei Sinica*, 2007, 28(4):115-118.
- [5] 杨世铭,陶文铨. 传热学[M]. 北京:高等教育出版社,2001.
- [6] 李克向. 保护油气层钻井完井技术[M]. 北京:石油工业出版社,1993.
- [7] BLAUER R E, MITCHELL B J. Determination of laminar, turbulent and transitional foam flow friction losses in pipes[R]. *SPE 4885*, 1974.

(编辑 沈玉英)