

光谱油样分析数据的神经网络多变量预测技术*

周 伽¹ 陈 果²

¹(南京航空航天大学理学院 南京 210016)

²(南京航空航天大学民航学院 南京 210016)

摘要 针对非等间隔的受加油因素影响的光谱油样分析数据的建模预测问题,利用神经网络高度的并行分布式、联想记忆能力、自组织学习能力和极强的非线性映射能力,建立了BP神经网络的多变量预测模型。同时,用遗传算法对网络参数进行了优化。最后,利用两组实际的航空发动机油样光谱分析数据对模型进行了验证。结果表明,提出的神经网络多变量预测模型能有效解决实际的受多因素影响的油样分析数据建模问题,具有较强的工程实用价值和通用性。

关键词 油样分析 多变量预测 非等间隔 加油 BP神经网络 遗传算法

中图分类号 O 329 F 201 **文献标识码** A **国家标准学科分类代码** 460.1540

Artificial Neural Network Multi-variable Forecasting Model for Spectral Oil Analysis Data

Zhou Jia¹ Chen Guo²

¹(College of Science, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

²(College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

Abstract The establishing model of the oil analysis data time series, which is unequal interval and affected by adding oil factor, is studied. As the artificial neural network (ANN) has the high parallel distributed processing, associative memory abilities, self-organization, self-learning and strong nonlinear mapping abilities, the multi-variable predicting model by back-propagation neural networks (BPNN) is established. In addition the genetic algorithm (GA) is used to optimize ANN's parameters. Finally, two time series of aero-engine oil spectral analysis data are used to verify this model. The result shows that this model can effectively solve the prediction problem of oil analysis data, which is affected by multi-factors such as unequal interval and adding oil. The new method has important engineering application value, and it is a common method of forecasting complex time series, which is affected by multi-factors.

Key words Oil analysis Multi-variable forecasting Unequal interval Add oil BP neural network Genetic algorithm (GA)

1 引 言

油样分析数据具有明显的非等间隔性。在机器的正常使用期故障率往往比较低,采样间隔比较大,而在损耗阶段故障率较高,这时往往要缩短采样间隔。由于采样时间间隔的不同形成了非等间隔时间序列。另外,

由于油样分析中,实际采样时间的安排本身也具有非等间隔性,因此难以进行完全等间隔采样。由此可见,非等间隔性具有普遍意义,等间隔时间序列只是其特殊情况。

同时,由于机器运行时有消耗滑油和出现滑油泄漏的情况,因此需要不定期的进行加油/补油或换油。由于每次取样后的加油影响,油样分析所得到的磨损

* 本文于 2004 年 12 月收到。

颗粒浓度将产生变化。这样就导致了实际油样分析数据将表现出受加油因素的影响特征。

目前,传统方法均不能有效地解决此类问题。传统ARMA法^[1]是针对平稳的、零均值、正态分布的随机序列的线性模型,因此不适合进行非线性预测;统计回归法进行非线性预测时需要知道序列的非线性模型,且回归阶次难于确定;灰色系统模型相当于指数回归模型,但是在实际应用中,一个非负序列的累加生成序列未必有指数规律,累加生成和累减还原的方法建模必将加大模型的误差,因此在实际应用中,要慎用灰色系统法^[2]。通常,当序列的趋势呈指数函数递增或递减时,灰色模型GM(1,1)具有相当高的精度,当系统发生转折或周期性变化时,其精度就变得相当差,灰色预测-校正模型^[3]尽管能弥补该缺陷,但需要对原始序列进行平滑,同时需要人工选取分段点,建模复杂,且通用性较差;神经网络方法是由于人工神经网络具有逼近任意非线性函数和一定的泛化能力,在预测领域显示出了强大的生命力和巨大的发展潜力。相关研究^[4-6]均表明了神经网络较其他方法的优越性,尤其是在多变量预测技术中,神经网络技术更加显示了其突出的优势。根据以上分析,精确预测航空发动机复杂磨损趋势,必须充分考虑多种因素(加油/补油或换油、非等间隔等),实施多变量预测方案。

这里旨在建立神经网络多变量预测模型,综合解决非等间隔的受加油因素影响的油样分析数据的建模和预测问题,从而为航空发动机的视情维修提供更为可靠的决策依据。

2 多层前向神经网络的反向传播(back propagation, BP)学习算法^[7]

BP算法将神经网络学习输入输出的映射问题变成一个非线性优化问题,使用了最优化中最普遍的梯度下降算法,用迭代运算修正网络权重,实现网络输出与期望输出间的均方误差最小化即目标函数 $E_p = \frac{1}{2} \sum_k (t_k - y_k)^2$ 最小化,其中, t 为实际网络输出, y 为期望输出。BP算法由前向计算过程和误差反向传播过程组成。在前向计算过程中,输入信息从输入层经隐层逐层处理(常用的非线性激活函数为Sigmoid函数),传向输出层(常用线性函数),每一层神经元的状态只影响下一层神经元的状态。如果在输出层得不到期望的输出,则转向反向传播,将误差信号沿原来的连接通路返回,利用权重的修正公式 $\Delta W_{kj} = -\eta \frac{\partial E_p}{\partial W_{kj}}$,通过修

改各层的权重 w 最终使得误差信号最小。

3 BP网络的多变量时间序列预测模型

3.1 神经网络多变量时间序列预测方法

对于多变量时间序列 $(x_{1i}, x_{2i}, \dots, x_{pi}), i = 1, 2, \dots, n$,它有 p 个时间变量,认为时间序列的未来值与其前面的 m 个值之间有某种函数关系:

$$\begin{aligned} (x_{1n+k}, x_{2n+k}, \dots, x_{pn+k}) = F &((x_{1n}, x_{2n}, \dots, x_{pn}), \\ (x_{1n-1}, x_{2n-1}, \dots, x_{pn-1}), \dots, &(x_{1n-m+1}, \\ x_{2n-m+1}, \dots, x_{pn-m+1})) \end{aligned} \quad (1)$$

利用神经网络拟合函数 F ,并用它进行预测。进行多变量时间预测的神经网络结构分为2种:单步预测网络和多步预测网络。单步预测网络输出的个数是 p ,即为1个多变量时间序列的变量个数,一次计算得到所有变量一步的预测值;多步预测网络的输出个数为 $k \times p$ 个,一次计算可得到所有 p 个变量的 k 步预测结果。多变量时间序列的多步预测实现方法是进行迭代多步预测。

利用传统的预测方法进行多变量时间序列预测的建模与预测非常复杂。而利用神经网络进行多变量时间序列的预测方法如同单变量时间序列预测一样简单,这是传统的预测方法无法比拟的。

3.2 油样分析数据的神经网络多变量预测模型

航空发动机光谱油样数据的非等间隔性是由于不定期取样和数据丢失等原因所致,因此,非等间隔的光谱油样数据是很常见的,通常的做法是将其近似作为等间隔处理,或通过插值得到等间隔数据,然后再运用通常的等间隔的数据建模方法(如ARMA模型)来对数据进行动态建模型预测方法。但这样的处理会造成较大的误差,与实际情况不符。有效的方法是直接对非等间隔数据进行建模,然后运用模型对未来短期内的任意时刻进行预测。对时间序列进行预测实际上是通过非线性系统产生的时间序列建模,把握非线性系统的运动规律和运动状态,通常的做法是从系统产生的时间序列中抽取动力系统,重构相空间,最常用的方法是时延法。

设所研究的时间序列为 $\{x(t), t = 1, 2, \dots, N\}$,则当前状态的信息可以表示成 m 维的延迟矢量:

$$x(t+k\tau) = f\{x(t), x(t-\tau), \dots, x(t-(m-1)\tau)\} \quad (2)$$

式中: m 为嵌入维数, τ 为时间延迟, k 为预测步数。通常取为采样间隔。Takens^[8]已经证明:假设动力系统的维数为 d ,如果 $m > 2d + 1$,则这种映射产生的伪相空

间和系统的状态空间微分同胚,及拓扑等价,他们的动力学特性定性意义上完全相同。

由此可见,对时间序列的预测,关键在于根据已知时间序列数据,对非线性系统相空间的重构,找出从 m 维空间映射到一维空间的映射函数。而神经网络用于预测,就是用其来拟合这一非线性函数。

对应式(2)可写为:

$$\begin{aligned}
 x(t+k\tau) &= f_1\{x(t), x(t-\tau), \dots, x(t-(m-1)\tau); \\
 &\quad t, t-\tau, \dots, t-(m-1)\tau, t+k\tau\} \\
 &= f_2\{x(t), x(t-\tau), \dots, x(t-(m-1)\tau, \\
 &\quad \tau, 2\tau, \dots, (m-1)\tau, k\tau\} \quad (3)
 \end{aligned}$$

对于非等间隔数据,令 $t = t + k\tau$, 则有:

$$\begin{aligned}
 x(t) &= f\{x(t_1), x(t_2), \dots, x(t_m); \\
 &\quad (t_2-t_1), (t_3-t_2), \dots, (t-t_m)\} \quad (4)
 \end{aligned}$$

由此可见,等间隔序列预测是非等间隔序列预测的特殊情况。下一 $k\tau$ 时刻的数值不仅与前 m 个数据的值有关,而且还与序列间的时间间隔有关,因此在建立非等间隔预测时,需要在神经网络的输入向量中考虑序列的时间间隔因素。

另外,由于航空发动机的滑油泄漏及受热蒸发,在每次取样之间均将消耗一部分滑油,为了保证机器充分润滑,每次取样后均要对滑油进行补充。而滑油的补充经常导致油样数据对每次滑油添加量的相关关系。下面对两者关系进行分析:

(1) 设 V (mL) 滑油中 Fe 元素的累计磨损量 (ng) 随时间 t (h) 的变化关系为 $W_{Fe}(t)$;

(2) 设换油量 (mL) 与时间 t 的关系为 $O_{\Delta}(t)$;

(3) i 时刻滑油中 Fe 元素浓度与时间的关系:

$$C_{Fei} = \left(W_{Fei} - \sum_{k=1}^{i-1} dW_{Fek} - C_{Fei} \times O_{\Delta i} \right) / V \quad (5)$$

式中: $C_{Fei} = \left(W_{Fei} - \sum_{k=1}^{i-1} dW_{Fek} \right) / V$

$$dW_{Fek} = C_{Fek} \times O_{\Delta k}$$

显然, i 时刻的 Fe 元素浓度不仅与当前时刻的 Fe 元素的累计磨损量 W_{Fei} 有关,还与前 $i-1$ 时刻的加油量有关。由时延法,式(2)可写为:

$$\begin{aligned}
 C(t+k\tau) &= f_1\{C(t), C(t-\tau), \dots, C(t-(m-1)\tau)\} \\
 &= f_2\{C(t), C(t-\tau), \dots, C(t-(\frac{t}{\tau})\tau); \\
 &\quad O_{\Delta}(t), O_{\Delta}(t-\tau), \dots, O_{\Delta}(t-(\frac{t}{\tau})\tau)\} \quad (6)
 \end{aligned}$$

由此可见,下一 τ 时刻的 Fe 元素浓度与以前时刻的浓度和加油量均有关系。

根据上述分析,文中建立综合考虑数据间隔信息、加油量信息的光谱油样数据的多变量神经网络预测模

型。为了简化模型,加油量的时间延迟与油样数据和采样时间均为 m 。如图 1 所示。

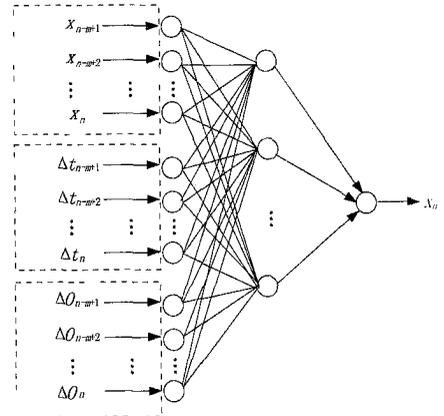


图 1 油样分析数据的神经网络多变量预测模型

(1) 将数据间的采样时间间隔 Δt 作为 1 个输入信息,从而形成 1 个既包含数据信息又包含时间间隔信息的训练样本和预测样本,这样使时间序列的间隔信息反映到了模型中。

(2) 将每次取样前的加油量 ΔO 作为 1 个变量,从而就形成 1 个既包含数据信息又包含加油信息的训练样本和预测样本。这样使加油信息反映到了模型中。

不失一般性,文中仅考虑单步预测情况,为统一解决非等间隔和受加油因素影响油样分析数据而建立的多变量预测模型是对单变量模型的推广,即需要拟合如式(7)所示的函数:

$$\begin{aligned}
 x_{n+1} &= F(x_{n-m+1}, x_{n-m+2}, \dots, x_n; \\
 &\quad \Delta t_{n-m+1}, \Delta t_{n-m+2}, \dots, \Delta t_n; \\
 &\quad \Delta O_{n-m+1}, \Delta O_{n-m+2}, \dots, \Delta O_n) \quad (7)
 \end{aligned}$$

式中: x 为数据信息, n 为样本序列号, m 为输入延时, Δt 为采样间隔, ΔO 为加油量。

4 遗传算法优化神经网络预测模型参数^[9]

神经网络预测能力的主要影响参数为输入层节点数 N 、中间隐层节点数 M 及网络允许训练误差 MSE , 为了获取最佳预测性能的 BP 网络,需要得到最佳的 N 、 M 和 MSE 目标值。显然这是一个优化问题,如果采取穷举的方式搜索最优值,计算量将十分巨大以至于无法实现。由于遗传算法具有隐含的并行性和强大全局搜索能力,可以在很短的时间内搜索到全局最优解。因此文中利用遗传算法来进行 BP 网络的预测模型的结构优化。

如图 2 所示,构造 BP 网络预测模型优化的遗传

算法需要以下步骤:

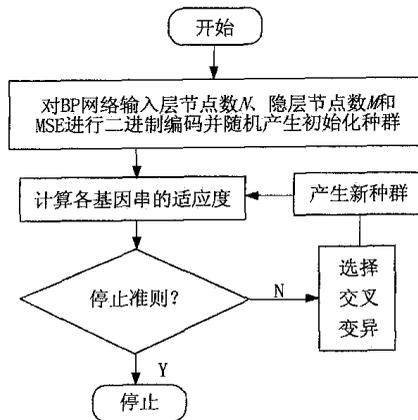


图2 神经网络结构优化的遗传算法流程

(1) 设定种群数目 n , 种群数目太小, 遗传算法的性能将变得很差或根本找不出问题的解, 太大, 则会增加计算量, 使收敛时间增长, 种群数目一般取为 30~50 个。

(2) 对输入层节点数 N 、中间隐层节点数 M 和 MSE 进行二进制编码, 并随机产生初始种群。根据优化参数的取值范围, 将其量化值(用二进制串表示)编码成基因串 $a = \{\alpha_0, \dots, \alpha_{A-1}; \alpha_A, \dots, \alpha_{A+B-1}, \alpha_{A+B}, \dots, \alpha_{A+B+C-1}\}$, 其中 a 中的前 A 个量化值代表输入层节点数 N , 中间 B 个量化值代表中间隐层节点数 M , 最后 C 个量化值代表 MSE 。所以每个基因串长度为 $A+B+C$ 个比特位, 此时的搜索空间有 2^{A+B+C} 个点。

(3) 适应度函数的确定。在实际应用中, 对于实际测得的时间序列 $\{x_1, x_2, \dots\}$, 可以利用其一部分数据建模, 而用另一部分数据来对所建模型进行验证, 如果预测值与实测值相差越少, 显然模型越理想, 理想情况是预测值与实测值相等, 则达到完美预测。通常衡量预测值与实测值差别的变量采用平均相对变动值 ARV (average relative variance), 其定义为:

$$ARV = \frac{\sum_{i=1}^N [x(i) - \hat{x}(i)]^2}{\sum_{i=1}^N [x(i) - \bar{x}(i)]^2} \quad (8)$$

式中: N 为比较数据个数, $x(i)$ 为实测数据值, \bar{x} 为实测数据平均值, $\hat{x}(i)$ 为预测值。显然, 平均相对变动值 ARV 越小, 表明预测效果越好, $ARV = 0$ 表示达到了理想预测效果; 当 $ARV = 1$ 时, 表明模型仅达到平均值的预测效果。由于 ARV 反映了网络的泛化能力, 所以定义适应度函数为:

$$f = \frac{1}{ARV} \quad (9)$$

(4) 对种群解码, 计算每条基因串的适应度。

(5) 将适应度最大的个体无条件地复制到下一代新种群中, 然后对父代种群进行选择、交叉和变异等遗传算子运算, 从而繁殖出下一代新种群其它 $n-1$ 个基因串。采用转轮法作为选取方法, 适应度大的基因串选择的机会大, 从而被遗传到下一代的机会大。交叉和变异是产生新个体的遗传算子, 交叉率太大, 将使高适应度的基因串结构很快被破坏掉, 太小则使搜索停止不前, 一般取 0.5~0.9。变异率太大, 将使遗传算法变为随机搜索, 太小则不会产生新个体, 一般取 0.01~0.1。

如果达到设定的繁衍代数, 返回最好的基因串, 并将其作为输入层节点数和隐层节点数, 算法结束。否则, 回到(4)继续下一代的繁衍。

5 应用实例

文中利用 MATLAB 中的神经网络工具箱^[10]对多变量预测模型进行编程实现。其中程序利用 `newff` 建立 3 层前向网络, 并用“`trainlm`”来对网络进行训练。

为了对文中提出的方法进行充分验证, 将分别用 2 组某型军用飞机发动机的非等间隔的受加油因素影响的滑油光谱分析的 Fe 元素含量数据对所建模型进行验证。该光谱数据是采用美国 BIRD 公司的原子发射光谱仪获得。在建模分析中, 均采用前面提出的神经网络多变量预测模型。

5.1 实例一

编号为 01z3221257 的航空发动机从 18~32h 的运行过程中, 得到了 66 个光谱油样分析数据, 将数据分成 2 个部分, 用其中的 90% 建模, 剩余的 10% 用于对所建模型进行外推验证。图 3 为加油量与发动机运行时间的关系, 从图中也可看出明显的非等间隔特征, 其中最小间隔为 1h, 最大为 10h。用遗传算法对神经网络进行优化得到, 在训练样本比例 0.9、允许训练步数 1000、隐层节点数 50、输入节点数 2、训练精度 0.001 下, 达到了最佳预测效果, 其预测值与实测值的比较结果如图 4 所示。从图中可以看出模型达到了相当高的预测精度。

5.2 实例二

编号为 01z4323005 的航空发动机从 3~345h 的运行过程中, 得到了 78 个光谱油样分析数据, 将采样数据分成两个部分, 用其中的 90% 建模, 剩余的 10% 用于对所建模型进行外推验证。图 5 为加油量与发动机运行时间的关系, 同时从图中也可看出明显的非等间隔特征, 最小间隔为 1h, 最大为 10h。用遗传算法对

神经网络进行优化得到,在训练样本比例 0.9、允许训练步数 1000、隐层节点数 50、输入节点数 2、训练精度 0.001 下,达到了最佳预测效果,其预测值与实测值的比较结果如图 6 所示。从图中可以看出模型达到了相当高的预测精度。

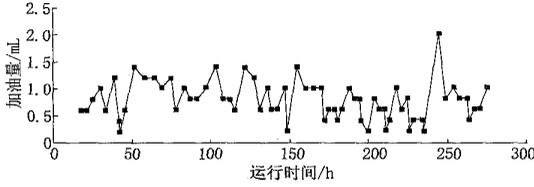


图 3 加油量与发动机运行时间的关系

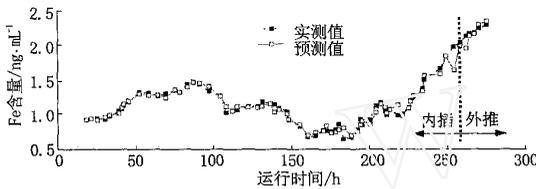


图 4 Fe 元素浓度与发动机运行时间的关系 (预测值与实测值的比较)

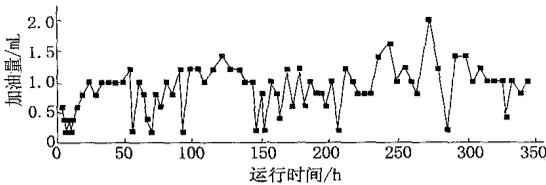


图 5 加油量与发动机运行时间的关系

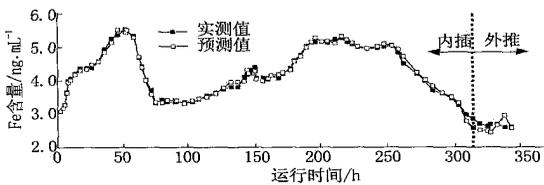


图 6 Fe 元素浓度与发动机运行时间的关系 (预测值与实测值的比较)

6 结 论

(1) 针对基于油样分析技术的航空发动机磨损状态趋势预测问题,提出了基于神经网络多变量预测模型的非等间隔的受加油因素影响的光谱油样分析数据的建模方法,以解决受不定期取样和加油等因素影响

的复杂时间序列预测问题

(2) 针对神经网络的结构设计困难问题,用遗传算法对神经网络预测模型参数——输入节点数、隐层节点数及网络收敛精度 3 个参数进行了优化。

(3) 利用实测的 2 组非等间隔的受加油因素影响的油样光谱数据验证了文中提出的神经网络多变量预测模型的有效性。

参考文献

[1] 杨叔子,吴雅 机械故障诊断的时序方法 西安:交通大学出版社,1989

[2] 张玉详,陆士良 神经网络在工程时间序列预报问题研究中的应用[J] 南京理工大学学报,1997,21(6): 522 - 524

[3] ZHANG H, L I Z G, CHEN Z N. Application of grey modeling method to fitting and forecasting wear trend of marine diesel engine [J] Tribology International, 2003(36): 753 - 756

[4] WEIGEND A B, et al Predicting the future: a connectionist approach [J] International Journal of Neural System, 1990(1): 193 - 209

[5] CHAKRABORTY K, et al Forecasting the behavior of multivariate time series using neural networks [J] Neural Networks, 1992, 5: 961 - 970

[6] 刘豹,胡代平 神经网络在预测中的一些应用研究 [J] 系统工程学报, 1999, 14(4): 338 - 344

[7] 吴今培,肖建华 智能故障诊断与专家系统 北京:科学出版社,1997: 1 - 139

[8] TAKENS F. Detecting strange attractors in turbulence //RAND D A, YOUNG L S Dynamical systems and turbulence Berlin: Springer-Verlag, 1981.

[9] 陈果 用遗传算法优化神经网络预测模型结构 中国机械工程[增刊], 2004: 176 - 179

[10] 飞思科技产品研发中心 MATLAB6.5 辅助神经网络分析与设计[M] 北京:电子工业出版社,2003

作者简介

周伽 女 1981 年出生 主要研究方向为神经网络遗传算法 应用数学

E-mail: zhou - jia81@126.com