

基于迭代容积卡尔曼滤波的神经网络训练算法

袁光耀¹ 胡振涛¹ 张 谨¹ 赵新强¹ 付春玲²

(河南大学图像处理与模式识别研究所 开封 475004)¹ (河南大学物理与电子学院 开封 475004)²

摘 要 针对现有应用非线性滤波算法对神经网络进行训练时存在精度不足的问题,提出了一种基于迭代容积卡尔曼滤波的神经网络训练算法。首先,将前馈神经网络各个节点的连接权值和偏置作为状态向量,建立前馈神经网络的状态空间模型。其次,利用 Spherical-Radial 准则生成容积点,并依据 Gauss-Newton 迭代策略来优化量测更新过程中获取的状态估计值和状态估计误差协方差,通过容积卡尔曼滤波估计精度的改善,提升神经网络节点的连接权值和偏置的训练效果。理论分析和仿真实验结果验证了所提算法的可行性和有效性。

关键词 前馈神经网络,状态空间模型,容积卡尔曼滤波,Gauss-Newton 迭代

中图法分类号 TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.10.048

Novel Neural Network Training Algorithm Based on Iterated Cubature Kalman Filter

YUAN Guang-yao¹ HU Zhen-tao¹ ZHANG Jin¹ ZHAO Xin-qiang¹ FU Chun-ling²

(Institute of Image Processing and Pattern Recognition, Henan University, Kaifeng 475004, China)¹

(School of Physics and Electronics, Henan University, Kaifeng 475004, China)²

Abstract In view of insufficient accuracy in the existing application of nonlinear filtering algorithm for neural network training, a novel neural network training algorithm based on iterated cubature Kalman filter was proposed. Firstly, the connection weights and bias of feedforward neural network are used as the state vector to establish the state space model. Secondly, using the Spherical-Radial standard to generate cubature points, the state estimation and covariance acquired during the measurement update process are optimized based on Gauss-Newton iteration strategy. The training effect of neural network connecting weights and bias is enhanced through the improvement of the estimation precision of cubature Kalman filter. The theoretical analysis and simulation results show the feasibility and effectiveness of the algorithm.

Keywords Feedforward neural network, State-space model, Cubature kalman filter, Gauss-Newton iterate

人工神经网络(ANN)具有大规模并行分布式结构特点,使其在并行计算、分布式存储、时序序列分析、模式识别、信号处理和控制等领域得到广泛应用。随着人工神经网络技术的发展,寻找高效的学习算法成为一个研究热点^[1]。在很长一段时间内,误差反向传播(Back Propagation, BP)方法成为神经网络训练中使用最广泛的一种算法。但是 BP 算法存在其固有缺陷,诸如易陷入局部极小值点、收敛速度缓慢、易受噪声和初始状态干扰等,极大限制了 BP 算法的使用领域和范围^[2]。随着基于贝叶斯估计的非线性滤波理论应用于神经网络的训练,神经网络的训练算法有了突破性的进展。Singhal 将基于非线性模型局部线性化思想的扩展卡尔曼滤波算法(EKF)应用于神经网络的训练,首先将神经网络各个节点的连接权值和偏置作为状态向量建立神经网络的状态空间模型,并将网络输出作为模型的观测值,然后用 EKF 算法更新神经网络的权值^[3]。该算法相比于 BP 算法显示出了极大的

优势,如收敛速度快、学习精度高、对噪声的适应性强,但是 EKF 算法也有其固有的缺陷,最致命的就是其所使用的局部线性化方法舍去了高阶项,易在强非线性情况下产生滤波发散。为了解决 EKF 算法存在的问题,文献[4]提出了一种基于无迹卡尔曼滤波(UKF)的神经网络训练算法,应用 UT 变换产生一组确定的加权采样点来获得神经网络状态向量的最优估计。与 EKF 算法相比,UKF 算法有效提高了神经网络的训练精度,但是当系统的状态向量高于三维时,为了避免状态协方差的非正定,需要凭借经验来调整 UKF 中的参数^[5]。de Freitas 和 Xi 将基于随机性采样策略的粒子滤波(PF)引入到神经网络的训练中,虽然提高了神经网络的训练精度,但其计算过程中所需的粒子数目较大,计算时间过长,难以在具体应用中推广^[6,7]。文献[8]提出了一种新的基于容积变换思想的状态估计方法——容积卡尔曼滤波算法(CKF),采用一组等权值的容积点来对后验概率密度进行近似。CKF 算法

到稿日期:2015-09-03 返修日期:2016-03-17 本文受国家自然科学基金(61300214),中国博士后基金(2014M551999),河南省高校科技创新团队支持计划(13IRTSTHN021),河南省高校青年骨干教师资助计划(2013GGJS-026),河南大学优秀青年培育基金(0000A40366),河南大学教学改革项目(2015)资助。

袁光耀(1989—),男,硕士生,主要研究方向为非线性滤波、智能计算;胡振涛(1979—),男,博士,副教授,主要研究方向为复杂系统建模与估计、多源信息融合;张 谨(1991—),女,硕士生,主要研究方向为目标跟踪、状态估计;赵新强(1979—),男,硕士,讲师,主要研究方向为神经网络、智能计算,E-mail:zhaoxq@henu.edu.cn(通信作者);付春玲(1980—),女,实验师,主要研究方向为多源信息融合、智能计算。

比 EKF 和 UKF 算法具备更高的滤波精度和滤波稳定性,并且无需选择任何参数,算法实现更为简便。此外,为了更好的滤波性能,Lefebvre 和 Mou 分别将量测迭代优化思想引入 EKF 和 UKF 算法,提出了迭代扩展卡尔曼滤波算法^[9]和迭代无迹卡尔曼滤波算法^[10],在量测更新过程中通过迭代过程充分利用最新的量测信息以提高滤波精度。

本文在 CKF 算法的基础上结合 Gauss-Newton 迭代策略更新过程中状态向量的均值和协方差,提出了一种基于迭代容积卡尔曼滤波(ICKF)的神经网络训练算法。该算法以前馈神经网络模型为基础,以前馈神经网络各个节点的连接权值和偏置量构成神经网络的状态空间向量,将网络的输出作为观测量建立神经网络的状态空间模型;随后应用 CKF 算法更新网络的权值和偏置量。当完成 CKF 算法的量测更新之后,引入迭代过程更新状态向量的均值和协方差从而得到更精确的网络权值和偏置的训练结果。仿真结果表明,应用 ICKF 算法对神经网络进行训练能得到更好的效果。

1 前馈神经网络的结构模型

典型的前馈神经网络(Feedforward Neural Network, FNN)由一个输入层、一个输出层和隐含层组成,如图 1 所示^[11]。层与层之间通过权值 W 连接, b 表示层之间的偏置量。神经网络的输入和输出之间的非线性映射关系如式(1)所示,式(1)中的上标表示层号, $h^1(\cdot)$ 和 $h^2(\cdot)$ 分别表示各层的非线性映射函数。

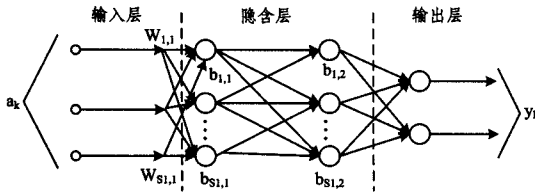


图 1 典型神经网络的模型结构

$$y_k = h^2(W^2 h^1(W^1 a_k + b^1) + b^2) + b^3 \quad (1)$$

将网络的所有连接权值和偏置节点组成神经网络的状态空间向量,则对于神经网络的训练就可以转化为通过 ICKF 算法来获得神经网络状态向量的最优估计。图 1 中的神经网络的状态转移和量测方程可以转化为如下形式:

$$w_{k+1} = w_k + q_k \quad (2)$$

$$y_k = h(w_k, a_k) + r_k \quad (3)$$

其中, w_k 表示网络中所有的连接权值和偏置节点组成的 k 时刻的状态向量, q_k 表示均值为零、协方差阵为 $\sigma_{q_k}^2$ 的过程噪声;第 $k+1$ 时刻的网络连接权值由 k 时刻的网络连接权值与系统的过程噪声所决定; $a_k \in R^d$ 表示神经网络的输入参数; r_k 表示均值为零、协方差阵为 $\sigma_{r_k}^2$ 的量测噪声; y_k 为网络的一步预测输出; $h(\cdot)$ 为神经网络的量测非线性映射,区别于式(1)中的 $h^1(\cdot)$ 和 $h^2(\cdot)$,此处 $h(\cdot)$ 为状态空间模型下整个神经网络的量测非线性映射^[11]。

2 迭代容积卡尔曼滤波算法

考虑一般离散非线性状态与量测方程:

$$x_{k+1} = f(x_k) + u_k \quad (4)$$

$$z_k = h(x_k) + v_k \quad (5)$$

其中, x_k 为系统的状态向量, u_k 表示过程噪声; $f(\cdot)$ 表示系

统的非线性状态转移方程; z_k 和 v_k 分别表示系统的量测值和量测噪声; $h(\cdot)$ 表示非线性量测方程; u_k 和 v_k 为相互独立的高斯白噪声序列,并且满足: $u_k \sim N(0, \sigma_{u_k}^2)$, $v_k \sim N(0, \sigma_{v_k}^2)$, $\sigma_{u_k}^2$ 和 $\sigma_{v_k}^2$ 分别为 u_k 和 v_k 的协方差阵。

标准的容积卡尔曼滤波算法利用 Spherical-Radial 准则,生成一组等权值的容积点,然后利用统计数值积分原理计算非线性变换后随机变量的均值和协方差。首先计算出各个容积点值以及相对应的权值^[8]。

$$\chi_i = \sqrt{\frac{m}{2}} \left\{ \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \dots, \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \dots, \begin{bmatrix} 0 \\ 0 \\ \vdots \\ -1 \end{bmatrix} \right\} \quad (6)$$

$$\omega_i = \frac{1}{m} \quad (7)$$

其中, m 表示容积点的个数,容积点的选取原则为状态向量维数的 2 倍,例如当状态空间向量的维数为 n_x 时, $m=2n_x$ 。 χ_i 表示生成的容积点; ω_i 表示每一个容积点所占的权值。在得到容积点之后,将这些点经过非线性系统的传递,可以得到较为精确的后验均值和协方差。随后结合卡尔曼滤波中状态预测与量测更新的思想实现对系统状态的估计。假设 k 时刻的状态协方差矩阵 P_k 正定,则令 $S_k = \text{chol}(P_k)$ 表示对协方差矩阵进行乔列斯基分解。迭代容积卡尔曼滤波算法的流程如下:

1) 时间更新

计算容积点 $X_{k,i}$ ($i=1,2,\dots,m$)。

$$X_{k,i} = S_k \chi_i + \hat{x}_k \quad (8)$$

其中, $S_k = \text{chol}(P_k)$, χ_i 表示应用式(6)生成的容积变换点, $\hat{x}_k$ 表示 k 时刻状态的最优估计。

将生成的容积点按照状态转移方程进行转移。

$$X_{k+1,i} = f(X_{k,i}) \quad (9)$$

利用状态转移之后的容积点计算状态 $\bar{x}_{k+1}$ 和协方差 $\bar{P}_{k+1}$ 。

$$\bar{x}_{k+1} = \sum_{i=1}^m \omega_i X_{k+1,i} \quad (10)$$

$$\bar{P}_{k+1} = \sum_{i=1}^m \omega_i X_{k+1,i} (X_{k+1,i} - \bar{x}_{k+1})^T + \bar{x}_{k+1} (\bar{x}_{k+1})^T + \sigma_{u_k}^2 \quad (11)$$

2) 量测更新

利用式(11)中生成的 $\bar{P}_{k+1}$ 重新进行乔列斯基分解,生成新的容积点 $\bar{X}_{k+1,i}$ 。

$$\bar{X}_{k+1,i} = \bar{S}_{k+1} \chi_i + \bar{x}_{k+1} \quad (12)$$

对新生成的容积点通过非线性方程进行量测。

$$Z_{k+1,i} = h(\bar{X}_{k+1,i}) \quad (13)$$

利用量测后的容积点得到量测的预测值 $\hat{Z}_{k+1}$,量测自协方差 P_{k+1}^{zz} 和互协方差 P_{k+1}^{xz} 以及卡尔曼滤波增益 K_{k+1} 。

$$\hat{Z}_{k+1} = \sum_{i=1}^m \omega_i Z_{k+1,i} \quad (14)$$

$$P_{k+1}^{zz} = \sum_{i=1}^m \omega_i Z_{k+1,i} (Z_{k+1,i} - \hat{Z}_{k+1})^T - \hat{Z}_{k+1} (\hat{Z}_{k+1})^T + \sigma_{v_k}^2 \quad (15)$$

$$P_{k+1}^{xz} = \sum_{i=1}^m \omega_i \bar{X}_{k+1,i} (Z_{k+1,i} - \hat{Z}_{k+1})^T - \bar{x}_{k+1} (\hat{Z}_{k+1})^T \quad (16)$$

$$K_{k+1} = P_{k+1}^{xz} (P_{k+1}^{zz})^{-1} \quad (17)$$

3) 状态和协方差更新

$$x_{k+1} = \bar{x}_{k+1} + K_{k+1} (z_{k+1} - \hat{z}_{k+1}) \quad (18)$$

$$P_{k+1} = \bar{P}_{k+1} - K_{k+1} P_{k+1}^{zz} (K_{k+1})^T \quad (19)$$

通过以上 3 个过程就完成了—次标准的容积卡尔曼滤波过程。

4) 迭代更新过程

由 CKF 算法的实现过程可知,在量测更新环节中 $k+1$ 时刻 x_{k+1} 的求解利用了状态—步预测结果的 $\bar{x}_{k+1}$ 和 $\bar{P}_{k+1}$,而量测更新环节中获取 x_{k+1} 和 P_{k+1} 值利用了 K_{k+1} 实现最新量测信息 z_{k+1} 对 $\bar{x}_{k+1}$ 和 $\bar{P}_{k+1}$ 的合理修正,其对于真实状态的逼近程度必然优于 $\bar{x}_{k+1}$ 和 $\bar{P}_{k+1}$ 。如果在量测更新环节中能够直接使用 x_{k+1} 和 P_{k+1} ,无疑将进一步提升对滤波器的估计精度。因此为了最大限度地利用已有的真实量测信息,结合 Gauss-Newton 迭代策略将迭代的方法引入状态的均值和协方差的更新之中^[13]。首先定义代价函数:

$$J(x_{k+1}) = (x_{k+1} - \bar{x}_{k+1})^T \bar{P}_{k+1}^{-1} (x_{k+1} - \bar{x}_{k+1}) + (y_{k+1} - h(\bar{x}_{k+1}))^T R_{k+1}^{-1} (y_{k+1} - h(\bar{x}_{k+1})) \quad (20)$$

应用 Gauss-Newton 迭代原理,可以得到 $J(x_{k+1})$ 取得极小值的迭代公式为(其中 j 表示第 j 次迭代):

$$x_{k+1}^{(j+1)} = \bar{x}_{k+1} + \bar{P}_{k+1} (H_{k+1}^{(j)})^T (H_{k+1}^{(j)} \bar{P}_{k+1} (H_{k+1}^{(j)})^T + R_{k+1})^{-1} \cdot [z_{k+1} - h(x_{k+1}^{(j)}) - H_{k+1}^{(j)} (\bar{x}_{k+1} - x_{k+1}^{(j)})] \quad (21)$$

其中, $H_{k+1}^{(j)} = \partial h(x_{k+1}) / \partial x_{k+1} |_{x_{k+1} = x_{k+1}^{(j)}}$ 。 $(H_{k+1}^{(j)} \bar{P}_{k+1} (H_{k+1}^{(j)})^T + R_{k+1})^{-1}$ 和 $\bar{P}_{k+1} (H_{k+1}^{(j)})^T$ 为量测方程线性化之后对于量测自协方差 $P_{k+1}^{zz(j)}$ 和状态量测之间的互协方差 $P_{k+1}^{xz(j)}$ 的近似,为了避免求偏导造成的舍入误差以及提高计算精度,利用第 j 次迭代过程中新生成的容积点 $\tilde{x}_{k+1,i}^{(j)}$ 和其对应的量测值 $Z_{k+1,i}^{(j)}$,采用式(15)—式(17)计算 $P_{k+1}^{zz(j)}$, $P_{k+1}^{xz(j)}$ 和 $K_{k+1}^{(j)}$,从而可以应用当前时刻的量测值迭代获得状态估计 $x_{k+1}^{(j+1)}$ 和协方差 $P_{k+1}^{(j+1)}$ 的更新。

$$x_{k+1}^{(j+1)} = \bar{x}_{k+1} + K_{k+1}^{(j)} [z_{k+1} - h(x_{k+1}^{(j)}) - (P_{k+1}^{xz(j)})^T \hat{P}_{k+1}^{-1} (\bar{x}_{k+1} - x_{k+1}^{(j)})] \quad (22)$$

$$P_{k+1}^{(j+1)} = \bar{P}_{k+1} - K_{k+1}^{(j)} (P_{k+1}^{zz(j)}) (K_{k+1}^{(j)})^T \quad (23)$$

ICKF 的算法流程图如图 2 所示。

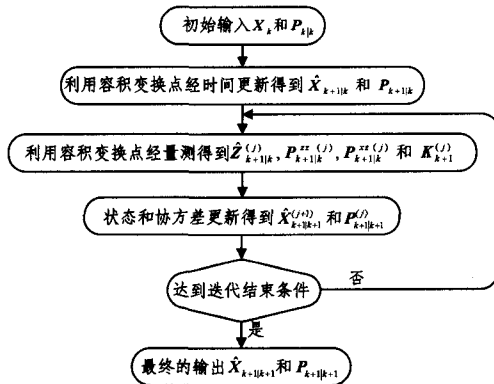


图 2 ICKF 算法流程图

3 基于 ICKF 的神经网络训练算法

文献[12]将 IUKF 算法应用于前馈神经网络的训练,并以此来预测 Mackey-Glass 时间序列。文献[14]将 IEKF 算法

应用于神经网络的训练并以此对引擎系统的可靠性进行了预测,显示了良好的训练效果。为了进一步提高神经网络的训练精度和训练效率,结合前馈神经网络模型,本节构建了一种基于 ICKF 的前馈神经网络训练算法(FNN-ICKF)。根据神经网络的状态空间模型和 ICKF 算法的执行流程,将基于 ICKF 的前馈神经网络训练算法(见图 3)的具体流程叙述如下。

步骤 1 根据式(4)和式(5)建立神经网络的状态空间模型。

步骤 2 初始化神经网络的权值和偏置节点,并作为容积卡尔曼滤波的初始状态 $\hat{w}_k = E(w_0)$,设置初始协方差矩阵 P_0 。

步骤 3 当 $k \geq 0$ 时,实现神经网络向前一步预测。

1) 进行时间更新,得到:

$$\bar{w}_{k+1} = \hat{w}_k \quad (24)$$

$$\bar{P}_{k+1} = P_k + \sigma_{q_k}^2 \quad (25)$$

2) 利用式(24)和式(25)得到的 $\bar{w}_{k+1}$ 和 $\bar{P}_{k+1}$ 进行容积变换,生成—组容积变换点 $\tilde{w}_{k+1,i}$ 。将 $k+1$ 时刻的输入值 a_{k+1} 输入神经网络中,按每一组容积变换点 $\tilde{w}_{k+1,i}$ 分别进行非线性映射,得到—组量测点 $Y_{k+1,i}$ 。

3) 利用生成的量测点 $Y_{k+1,i}$ 计算得出量测的预测值 $\hat{Y}_{k+1}$,量测自协方差 P_{k+1}^{yy} 和状态与量测互协方差 P_{k+1}^{wy} 以及卡尔曼滤波增益 K_{k+1} :

$$\hat{Y}_{k+1} = \sum_{i=1}^m \omega_i Y_{k+1,i} \quad (26)$$

$$P_{k+1}^{yy} = \sum_{i=1}^m \omega_i Y_{k+1,i} (Y_{k+1,i})^T - \hat{Y}_{k+1} (\hat{Y}_{k+1})^T + \sigma_{v_k}^2 \quad (27)$$

$$P_{k+1}^{wy} = \sum_{i=1}^m \omega_i \tilde{w}_{k+1,i} (Y_{k+1,i})^T - \bar{w}_{k+1} (\hat{Y}_{k+1})^T \quad (28)$$

$$K_{k+1} = P_{k+1}^{wy} (P_{k+1}^{yy})^{-1} \quad (29)$$

4) 获得 $k+1$ 时刻真实量测值 y_{k+1} 之后,更新状态和协方差:

$$w_{k+1} = \bar{w}_{k+1} + K_{k+1} (y_{k+1} - \hat{Y}_{k+1}) \quad (30)$$

$$P_{k+1} = \bar{P}_{k+1} - K_{k+1} P_{k+1}^{zz} (K_{k+1})^T \quad (31)$$

5) 当步骤 4) 进行完毕之后就完成了一次标准的 CKF 算法训练神经网络的流程。如果能充分利用最新的量测信息 y_{k+1} 对 w_{k+1} 和 P_{k+1} 进行合理的修正,其对于真实状态的逼近程度必然更优。因此,为了提高算法的精度,将迭代更新的过程引入神经网络的训练之中。利用步骤 1) 中得到的 $\bar{w}_{k+1}$ 和 $\bar{P}_{k+1}$ 作为初始的迭代过程。

FOR $j=1:N_{\max}$ (其中 j 表示迭代次数, $N_{\max}$ 表示最大迭代次数)

— 产生新的容积变换点 $\tilde{w}_{k+1,i}^{(j)}$

— 经过网络的映射得到输出预测值 $\hat{Y}_{k+1}^{(j)}$, 自协方差 $P_{k+1}^{yy(j)}$ 和互协方差 $P_{k+1}^{wy(j)}$ 以及卡尔曼滤波增益 $K_{k+1}^{(j)}$

— 状态和协方差更新

$$w_{k+1}^{(j+1)} = \bar{w}_{k+1} + K_{k+1}^{(j)} [y_k - h(w_{k+1}^{(j)}, a_{k+1}) - (P_{k+1}^{wy(j)})^T \hat{P}_{k+1}^{-1} (\bar{w}_{k+1} - w_{k+1}^{(j)})] \quad (32)$$

$$P_{k+1}^{(j+1)} = \bar{P}_{k+1} - K_{k+1}^{(j)} (P_{k+1}^{zz(j)}) (K_{k+1}^{(j)})^T \quad (33)$$

— 当 $\|w_{k+1}^{(j+1)} - w_{k+1}^{(j)}\| \leq$ 误差阈值或达到最大迭代次

数,结束迭代过程。否则继续下一轮迭代过程。

END

步骤4 令 $k=k+1$,跳转到步骤1)。

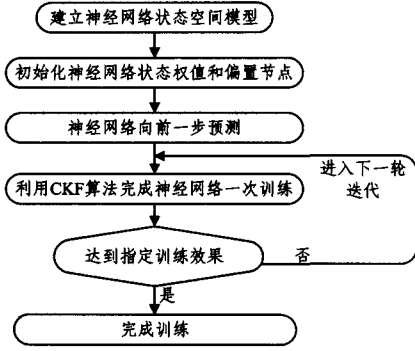


图3 基于 ICKF 的神经网络训练算法

4 仿真实验与分析

为验证算法的可行性与有效性,选取前馈神经网络(FNN)作为模型。仿真实验算例中分别采用了单输入单输出模型和多输入多输出模型。考虑到算法的鲁棒性能,仿真实验采用 Monte Carlo 仿真形式,仿真次数取 50,并以均方根误差(RMSE)作为算法精度的衡量指标。分别将使用 IEKF 算法和 IUKF 算法对前馈神经网络的训练记为 FNN-IEKF 和 FNN-IUKF 作为对比算法,以 3 种算法单步运行所用时间作为算法耗时的评价依据。仿真环境为 Intel i5 CPU,主频 2.5GHz,4GB 内存,仿真软件采用 Matlab R2012a。

算例1 单输入单输出模型

$$x_k = 0.5x_{k-1} + \frac{25x_{k-1}}{(1+x_{k-1}^2)} + 8\cos(1.2(k-1)) + u_k \quad (34)$$

$$y_k = \frac{x_k^2}{20} + 6\sqrt{0.05(k-1)} + 3 + v_k \quad (35)$$

其中, x_k 表示输入向量, y_k 表示输出向量,仿真步长为 120,过程噪声 u_k 为均值为 0、方差为 0.01 的高斯白噪声,量测噪声均值为 0,方差为 3。系统的初始输入 $x_0 = 0.1$ 。前馈神经网络的结构如图 1 所示,由于是单输入单输出系统,输入层和输出层的节点数目均为 1,隐含层节点数目设为 10 个,感知器选择 sigmoid 型函数,输出层为线性输出。神经网络的初始权值为方差为 10 的一组随机数,ICKF 的初始状态协方差 P 、状态转移噪声协方差 σ_k^2 和量测噪声协方差 σ_k^2 分别设为 1, $1E-5$, 1。迭代次数均选择为 3 次,分别进行 50 次 Monte Carlo 仿真来验证算法的稳定性。

图 4 给出了 3 种算法的向前一步预测结果比较,表 1 给出了 3 种算法平均一步状态预测的均方根误差(RMSE)及平均一步预测运行时间的结果比较。从理论上分析,执行一次标准 EKF、UKF 以及 CKF 算法的计算复杂度分别为^[15]:

$$f_{EKF} = 2n^3 + n^2 + 2n^2l + 2nl^2 - nl - n + l \quad (36)$$

$$f_{UKF} = \frac{26n^3}{3} + 15n^2 + 10n^2l + 5n + 8nl^2 + 6nl + l^3 + 3l^2 + 4l \quad (37)$$

$$f_{CKF} = \frac{20n^3}{3} + 10n^2 + 10n^2l + 8nl^2 + 2nl + l^3 + 3l^2 + l \quad (38)$$

其中, n 表示状态的维数, l 表示量测维数。引入迭代过程之后,3 种算法的复杂度则与具体的迭代次数成倍数关系。从式(36)一式(38)中可以看出,EKF、UKF 和 CKF 3 种算法的

复杂度均为 $O(n^3)$ 量级,并且存在 $f_{UKF} > f_{CKF} > f_{EKF}$ 的关系。从仿真结果可以看出:使用 ICKF 算法的训练精度优于 IEKF 和 IUKF 算法,并且运算时间相较于 IUKF 算法也略有降低,这与算法耗时的理论分析也相一致。图 5 给出了不同迭代次数下的误差比较,从图中可以看出在迭代次数达到 3 次以后各种算法的表现趋于稳定,因此在实际使用时选择 3 次迭代即可。

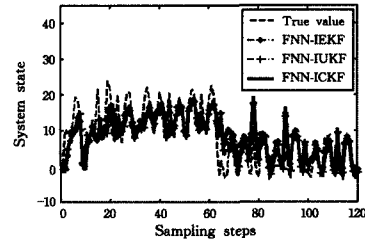


图4 系统状态估计

表1 平均一步预测 RMSE 与算法耗时

算法	RMSE	耗时(s)
FNN-IEKF	4.437	0.058
FNN-IUKF	4.327	0.325
FNN-ICKF	4.108	0.312

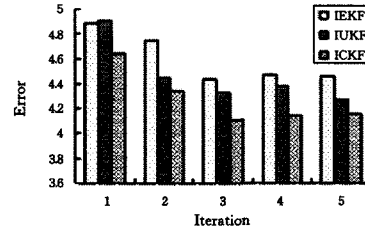


图5 迭代次数与误差关系

算例2 多输入多输出模型

设定仿真场景为二维平面内实现对匀速转弯运动目标的一步预测追踪,目标系统状态方程为:

$$x_{k+1} = Fx_k + \Gamma u_k \quad (39)$$

其中,目标的状态向量表示为 $x_k = [x_k \dot{x}_k y_k \dot{y}_k]^T$; 状态转移

$$\text{矩阵 } F \text{ 为 } \begin{bmatrix} 1 & \frac{\sin(\omega T)}{\omega} & 0 & \frac{-(1-\cos(\omega T))}{\omega} \\ 0 & \cos(\omega T) & 0 & -\sin(\omega T) \\ 0 & \frac{(1-\cos(\omega T))}{\omega} & 1 & \frac{\sin(\omega T)}{\omega} \\ 0 & \sin(\omega T) & 0 & \cos(\omega T) \end{bmatrix}, \text{ 状态}$$

向量中 x_k 和 y_k 表示 x 轴和 y 轴上的位置, $\dot{x}_k$ 和 $\dot{y}_k$ 表示 x 轴和 y 轴方向上的速度分量。 ω 为目标的转弯角速度,取 $\omega = 0.04 \text{ rad/s}$ 。 T 为采样间隔,在此取 $T=1$,过程噪声 u_k 为高斯白噪声,方差为 $\sigma_k^2 = \text{diag}[1 \ 1]$ 。噪声转移矩阵为 $\Gamma = \begin{bmatrix} T^{\wedge} 2/2 & T & 0 & 0 \\ 0 & 0 & T^{\wedge} 2/2 & T \end{bmatrix}^T$ 。雷达对于目标的量测分别为径向距 r_k 和方位角 θ_k 。量测方程如下:

$$\begin{bmatrix} r_k \\ \theta_k \end{bmatrix} = \begin{bmatrix} \sqrt{x_k^2 + y_k^2} \\ \tan^{-1}(\frac{y_k}{x_k}) \end{bmatrix} + v_k \quad (40)$$

v_k 为高斯白噪声序列,其方差 $\sigma_k^2 = \text{diag}[25 \ 0.01]$ 。令目标的初始状态为 $x_0 = [300 \ 30 \ 300 \ 10]^T$ 。网络参数除输入层与输出层的节点数均为 2 外,其余参数的选取同算例 1,

网络的输入参数为水平方向的位置 x_k 和垂直方向的位置 y_k , 输出为径向距 r_k 和方位角 θ_k 的一步预测。Monte Carlo 仿真次数为 50 次, 仿真步长为 120。

图 6 和图 7 分别给出了 3 种算法对于径向距和方位角的一步预测结果比较, 图 8 和图 9 给出了 50 次独立实验条件下各种算法的均方根误差 (RMSE) 的比较。平均一步预测 RMSE 与算法耗时如表 2 所列。从仿真结果可以看出, IUKF 和 ICKF 算法的收敛速度明显快于 IEKF 算法, 并且当目标的运动状态发生突变时能更快地进行相应的调整。ICKF 算法相比于 IUKF 算法在没有明显增加耗时的前提下精度上也有所提高, 并且 ICKF 算法的参数选取有严格的理论依据, 不会出现 IUKF 算法中参数选取不当易发生协方差非正定的情况, 这一点虽无法在仿真中有所体现, 但是在实际应用中却更加简便易行。

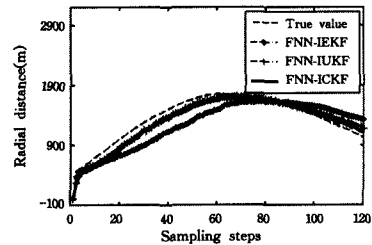


图 6 径向距状态估计

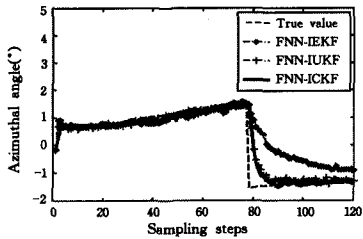


图 7 方位角状态估计

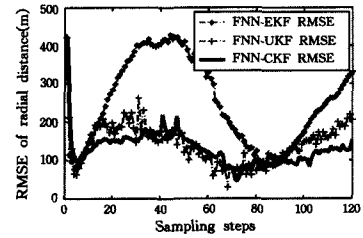


图 8 径向距状态估计 RMSE

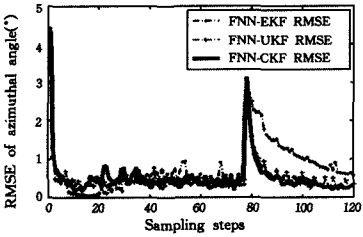


图 9 方位角状态估计 RMSE

表 2 平均一步预测 RMSE 与算法耗时

算法	径向距 (m)	方位角 (°)	耗时 (s)
FNN-IEKF	2.504E+02	0.746	0.207
FNN-IUKF	1.420E+02	0.541	0.882
FNN-ICKF	1.286E+02	0.510	0.856

结束语 近年来,随着神经网络在并行计算和大数据理

论上的应用,神经网络的应用范围进一步拓展。对于系统模型的大型化复杂化,已有的对于神经网络的训练算法的效率变得低下且学习效果不佳。在深入分析现有训练算法存在的缺点与不足的前提下,本文提出了一种新的基于迭代容积卡尔曼滤波的神经网络训练算法,有效弥补了现有算法的不足,提高了神经网络的训练精度。根据 ICKF 算法的构建过程和仿真结果分析,相对于现有的神经网络训练算法,基于 ICKF 的神经网络训练算法具有以下优势:首先,CKF 算法避免了 EKF 算法中局部线性化带来的高阶舍入误差问题,而且比 UKF 算法具备更高的滤波精度和稳定性,参数选取简便,算法更易于实现。其次,为了最大限度地利用已有的真实量测信息,将 Gauss-Newton 迭代策略引入状态的均值和协方差的更新之中,进一步提高了 CKF 算法的精度。需要说明的是,由于引入了迭代的过程,致使整个算法的实时性指标有所降低,如何更进一步地优化 ICKF 参数的选取、提高算法的运行效率将成为下一步研究的重点。

参考文献

- [1] Li Yuan, Deng Hui-wen. Application of speech recognition system based on algebra algorithm [J]. Computer Science, 2008, 35 (8): 122-124 (in Chinese)
李原, 邓辉文. 新型前馈网络学习算法在语音识别中的应用 [J]. 计算机科学, 2008, 35 (8): 122-124
- [2] Arsiy E A, Antoshchuk S G. Improving the efficiency of MLP back propogation learning at the classification of quasi-stationary signals [C] // Intelligent Data Acquisition and Advanced Computing Systems (IDAACS). 2011: 365-368
- [3] Singhal S, Wu L. Training multilayer perceptrons with the extended Kalman algorithm [C] // Advance in Neural Information Processing Systems, California: Morgan Kaufman Publishers Inc. 1989: 133-140
- [4] Wu X D, Wang Y N. Extended and Unscented Kalman filtering based feedforward neural networks for time series prediction [J]. Applied Mathematical Modelling, 2012, 36 (3): 1123-1131
- [5] Zarei J, Shokri E. Robust fault detection via unknown input observer based on cubature Kalman filter [J]. 21st Iranian Conference on Electrical Engineering (ICEE), 2013, 28 (2): 303-308
- [6] Xi Y H, Peng H. MLP training in a self-organizing state space model using unscented Kalman particle filter [J]. IEEE Transactions on Systems Engineering and Electronics, 2013, 24 (1): 141-146
- [7] de Freitas J F G, Niranjan M, Gee A H, et al. Sequential monte catlo methods to train neural network models [R]. England: Cambridge University, 1998
- [8] Arasaratnam I, Haykin S, Hurd T R. Cubature Kalman Filtering for continuous-discrete system: theory and simulations [J]. IEEE Transactions on Signal Processing, 2010, 58 (10): 4977-4993
- [9] Lefebvre T, Bruyninckx H, Schutter J D. Kalman filters for non-linear systems: a comparison of performance [J]. International Journal of Control, 2004, 77 (7): 639-653
- [10] Mou Z K, Sui L F. Improvement of UKF Algorithm and Robustness Study [C] // International Workshop on Intelligent Systems and Applications. 2009: 1-4
- [11] Zhang Yi-peng, Chen Liang, Hao Huan. LM based training algorithm for quantum neural networks [J]. Computer Science,

2013,40(9):221-224(in Chinese)

张翼鹏,陈亮,郝欢. 一种基于 LM 的量子神经网络训练算法[J]. 计算机科学,2013,40(9):221-224

- [12] Zha F,Xu J N,Li J S,et al. IUKF neural network modeling for FOG temperature drift[J]. Journal of Systems Engineering and Electronics,2013,24(5):828-844
- [13] Mu J,Cai Y L. Iterated cubature Kalman filter and its application[C]//IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER). Kunming,China,2011:33-37

(上接第 241 页)

- [2] Park J,Kim B I. The school bus routing problem: A review [J]. European Journal of Operational Research,2010,202(2):311-319
- [3] Dang Lan-xue,Chen Xiao-pan,Kong Yun-feng. Reivew of School Bus Routing Problem: Concept, Model And Optimizaion Algorithms[J]. Journal of Henan University(Natutal science),2013,43(6):682-691(in Chinese)
党兰学,陈小潘,孔云峰. 校车路径问题模型及算法研究进展[J]. 河南大学学报(自然科学版),2013,43(6):682-691
- [4] Dang Lan-xue,Wang Zhen,Liu Qing-song,et al. Heuristic Algorithm for Solving Mixed Load School Bus Routing Problem [J]. Computer Science,2013,40(7):248-253(in Chinese)
党兰学,王震,刘青松,等. 一种求解混载校车路径的启发式算法 [J]. 计算机科学,2013,40(7):248-253
- [5] Dror M, Trudeau P. Savings by split delivery routing [J]. Transportation Science,1989,23(2):141-145
- [6] Dror M, Trudeau P. Split delivery routing [J]. Naval Research Logistics (NRL),1990,37(3):383-402
- [7] Archetti C,Savelsbergh M W P,Grazia Speranza M. To split or not to split: That is the question [J]. Transportation Research Part E: Logistics and Transportation Review,2008,44(1):114-123
- [8] Silva M M,Subramanian A,Ochi L S. An iterated local search heuristic for the split delivery vehicle routing problem [J]. Computers & Operations Research,2015,53(53):234-249
- [9] Frizzell P W, Giffin J W. The split delivery vehicle scheduling problem with time windows and grid network distances [J]. Computers & Operations Research,1995,22(6):655-667
- [10] Archetti C,Speranza M G,Hertz A. A tabu search algorithm for the split delivery vehicle routing problem [J]. Transportation Science,2006,40(1):64-73
- [11] Berbotto L,García S,Nogales F J. A Randomized Granular Tabu Search heuristic for the split delivery vehicle routing problem [J]. Annals of Operations Research,2014,222(1):153-173
- [12] Liu Wang-sheng,Yang Fan,Li Mao-qing,et al. Clustering algorithm for split delivery vehicle routing problem[J]. Control and Decision,2012,27(4):535-541(in Chinese)
刘旺盛,杨帆,李茂青,等. 需求可拆分车辆路径问题的聚类求解算法 [J]. 控制与决策,2012,27(4):535-541
- [13] Meng Fan-chao,Lu Zhi-qiang,Sun Xiao-ming. Taboo Search algorithm of split Vehicle routing problem [J]. Computer Aided Engineering,2010,19(1):78-83(in Chinese)
孟凡超,陆志强,孙小明. 需求可拆分车辆路径问题的禁忌搜索算法 [J]. 计算机辅助工程,2010,19(1):78-83

- [14] Wu X D,Chang Y C,Mao J X,et al. Predicting reliability and failures of engine systems by single multiplicative neuron model with iterated nonlinear filters[J]. Reliability Engineering and System Safety,2013,119:244-250
- [15] Zhang Shao-you,Hao Yan-ling,Wu Xu. Complexity analysis of three deterministic sampling nonlinear filtering algorithms[J]. Journal of HARBIN Institute of Technology,2013,45(12):111-115(in Chinese)
张召友,好烟令,吴旭. 3 种确定性采样非线性滤波算法的复杂度分析[J]. 哈尔滨工程大学学报,2013,45(12):111-115
- [14] Ho S C,Haugland D. A tabu search heuristic for the vehicle routing problem with time windows and split deliveries [J]. Computers & Operations Research,2004,31(12):1947-1964
- [15] Belfiore P,Yoshizaki H T Y. Heuristic methods for the fleet size and mix vehicle routing problem with time windows and split deliveries [J]. Computers & Industrial Engineering,2013,64(2):589-601
- [16] Belfiore P C,Yoshida Yoshizaki H T. Scatter search for a real-life heterogeneous fleet vehicle routing problem with time windows and split deliveries in Brazil [J]. European Journal of Operational Research,2009,199(3):750-758
- [17] Thangiah S R,Fergany A,Wilson B,et al. School bus routing in rural school districts [M]. Computer-aided Systems in Public Transport. Springer,2008:209-232
- [18] Jin Yan-bo. Research on Optimizatin Of School Bus Routing [D]. Changchun:Jilin University,2006(in Chinese)
金燕波. 校车路径优化问题研究 [D]. 长春:吉林大学,2006
- [19] Archetti C, Mansini R, Speranza M G. Complexity and Reducibility of the Skip Delivery Problem [J]. Transportation Science,2005,39(2):182-187
- [20] Archetti C,Speranza M G,Savelsbergh M W P. An Optimization-Based Heuristic for the Split Delivery Vehicle Routing Problem [J]. Transportation Science,2008,42(1):22-31
- [21] Lourenço H,Martin O,Stützle T. Iterated Local Search: Framework and Applications(2nd ed)[M]//Gendreau M,Potvin J-Y. Handbook of Metaheuristics. Springer US,2010:363-397
- [22] Schrimpf G,Schneider J,Stamm-Wilbrandt H,et al. Record Breaking Optimization Results Using the Ruin and Recreate Principle [J]. Journal of Computational Physics,2000,159(2):139-171
- [23] Schneider J,Kirkpatrick S. Ruin & Recreate [M]//Schneider J, Kirkpatrick S. Stochastic Optimization. Berlin: Springer Berlin Heidelberg,2006:73-77
- [24] Bent R, Van Hentenryck P. A Two-Stage Hybrid Local Search for the Vehicle Routing Problem with Time Windows [J]. Transportation Science,2004,38(4):515-530
- [25] Bent R,Hentenryck P V. A two-stage hybrid algorithm for pick-up and delivery vehicle routing problems with time windows [J]. Computers & Operations Research,2006,33(4):875-893
- [26] Park J,Tae H,Kim B-I. A post-improvement procedure for the mixed load school bus routing problem [J]. European Journal of Operational Research,2012,217(1):204-213
- [27] Braca J,Bramel J,Posner B,et al. A computerized approach to the New York City school bus routing problem [J]. IIE Transactions,1997,29(8):693-702