

改进深度信念网络在飞机下降段油耗估计中的应用

刘家学 尹鹏

(中国民航大学电子信息与自动化学院 天津 300300)

摘要 为准确估计飞机油耗量,针对飞机下降段油耗影响因素众多且与油耗呈非线性关系的特点,提出一种基于改进深度信念网络(DBN)模型的飞机下降阶段油耗估计方法。通过引入高斯-伯努利受限玻尔兹曼机(GBRBM),解决传统DBN模型中受限玻尔兹曼机(RBM)处理连续油耗输入数据时信息丢失问题;采用自适应步长(AS)策略加快收敛速度。该方法利用改进DBN模型的深层非线性网络结构实现油耗影响因素与油耗复杂非线性函数关系的逼近,并通过模型顶层连接的BP网络进行油耗估计。实验结果表明,改进DBN模型在复杂非线性估计方面有较大优势,油耗估计精度和拟合度较高,验证了该方法在飞机下降段油耗估计领域具有可行性。

关键词 下降段油耗估计 油耗影响因素 改进深度信念网络 高斯-伯努利受限玻尔兹曼机 自适应步长

中图分类号 TP391.9

文献标识码 A

DOI:10.3969/j.issn.1000-386x.2019.08.013

APPLICATION OF IMPROVED DBN IN FUEL CONSUMPTION ESTIMATION OF AIRCRAFT DOWN SECTION

Liu Jiaxue Yin Peng

(College of Electronic Information and Automation, Civil Aviation University of China, Tianjin 300300, China)

Abstract In order to accurately estimate the fuel consumption of aircraft, aiming at the many factors affecting the fuel consumption of the descending section of the aircraft and its non-linear relationship with fuel consumption, this paper proposed a fuel consumption estimation method based on the improved deep belief network (DBN) model. We solved the problem of information loss when the restricted Boltzmann machine (RBM) processed the continuous fuel consumption input data in the traditional DBN model by introducing the Gauss-Bernoulli restricted Boltzmann machine (GBRBM). The adaptive step (AS) strategy was used to speed up convergence. This method used the deep nonlinear network structure of the improved DBN model to achieve the approximation of the complex nonlinear function relationship between fuel consumption factors and fuel consumption, and the fuel consumption was estimated by BP network connected at the top of the model. The experimental results show that the improved DBN model has great advantages in complex nonlinear estimation, and the fuel consumption estimation accuracy and fitting degree are high. It is proved that the method is feasible in the fuel consumption estimation of aircraft descending section.

Keywords Fuel consumption estimation of down section Influence factor of fuel consumption Improved deep belief network Gauss-Bernoulli restricted boltzmann machine Adaptive step

0 引言

随着航空运输的不断发展,航空煤油的消耗量逐年增长,民航业在能源、环境方面面临着越来越大的挑战。对民航业而言,精细化燃油管理有助于减少燃油

消耗,因此设计良好的油耗模型对提高燃油利用率减少民航碳排放具有重大意义^[1-2]。

飞机飞行过程主要包括爬升、巡航和下降三个阶段,这三个阶段的油耗量占整个飞行过程的90%以上,当航程越短时,下降段油耗量占比越大。许多学者在飞机油耗估计领域做了深入研究,并提出多种方法

进行建模分析,如线性回归模型^[3-4]、模糊逻辑控制模型^[5]、支持向量回归模型^[6]等。这些方法虽然不同程度地提高了估计精度,但由于飞机下降段油耗会受到气象条件和飞机性能衰减的影响,估计精度和鲁棒性并不高。为了解决以上问题,文献[7]提出了基于 BP 神经网络的油耗估计模型,利用神经网络良好的抗干扰和非线性映射能力,提高了估计精度和鲁棒性;文献[8]利用遗传算法对 BP 神经网络模型进行了优化,进一步提高了模型性能和油耗估计精度。然而飞机下降段油耗影响因素众多,非线性关系复杂,BP 神经网络模型属于浅层模型,对复杂函数的表达能力有限。深度信念网络(DBN)^[9]模型由多个受限玻尔兹曼机(RBM)堆叠而成,具有深层非线性网络结构,可以用来解决复杂非线性问题,但是 RBM 可见层的输入被限制为二值^[10],当处理连续油耗输入数据时会导致信息丢失,影响模型估计精度。

针对现有方法存在的不足,本文提出了一种基于改进 DBN 模型的下降段油耗估计方法。通过引入高斯-伯努利受限玻尔兹曼机(GBRBM),解决受限玻尔兹曼机(RBM)在处理连续油耗输入数据时信息丢失问题;采用自适应步长(AS)减少 DBN 训练时间。本文利用改进 DBN 模型的深层非线性网络结构^[11]来实现油耗影响因素与油耗之间复杂非线性函数关系的逼近。先通过多层 RBM 对油耗影响因素输入向量做逐层非线性变换,再利用顶层的 BP 网络对变换后的数据进行油耗估计,最后得到油耗估计值。文中以均方根误差、平均绝对百分比误差以及拟合优度为评价基准,判断模型的实用性。

1 下降段油耗影响因素分析及选取

飞机快速存储记录器(Quick Access Recorder, QAR)中记录了许多下降段油耗影响因素数据,包括飞机重量、下降距离、下降率、风速等诸多参数。各影响因素对油耗的影响程度不尽相同,以下降段油耗与下降距离关系为例,随着下降距离的增加飞机油耗不断增多,下降距离对下降段油耗影响显著,且下降距离与油耗之间呈非线性关系。

本文根据 QAR 数据中记录的油耗影响因素数据,逐一分析与油耗的关系,最终选取飞机总重、下降距离、下降率、风速、风向、大气总温、俯仰角、迎角、左发转速、右发转速等 10 个参数作为油耗估计模型输入。各参数与油耗呈非线性关系且相比之前方法模型参数增多导致模型更加复杂,本文利用改进的深度信念网络(GBRBM-DBN)模型解决这些问题。

2 GBRBM-DBN 油耗估计模型

Hinton 等在 2006 年提出深度学习的概念。深度信念网络(DBN)是深度学习中常用的一种网络模型,它是基于能量概率分布的模型,由多个 RBM 模型堆叠而成,上一层(低层)RBM 的隐藏层(输出层)作为下一层(高层)RBM 的可见层(输入层),每一层训练后的输出都是对输入的非线性变换。GBRBM-DBN 油耗估计模型利用高斯-伯努利受限玻尔兹曼机(GBRBM)替代第一层 RBM,处理连续的油耗输入数据。

2.1 受限玻尔兹曼机

受限玻尔兹曼机(RBM)是一个两层的无向图模型,包含一个可见层 V 和一个隐藏层 H ,如图 1 所示。可见层和隐藏层的神经元在层内无连接,层间全连接。可见层用于接收输入数据,隐藏层对可见层进行重构。

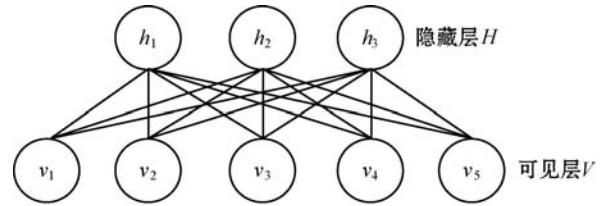


图 1 RBM 结构

假设所有的可见层单元 v 和隐藏层单元 h 均为二值分布,即 $\forall i, j, v_i \in \{0, 1\}, h_j \in \{0, 1\}$ 。对于给定的一组状态 (v, h) ,RBM 的能量函数为:

$$E_{\theta}(v, h) = - \sum_{i=1}^n a_i v_i - \sum_{j=1}^m b_j h_j - \sum_{i=1}^n \sum_{j=1}^m v_i w_{ij} h_j \quad (1)$$

式中: n 为可见层单元数目; m 为隐藏层的单元数目; $\theta = \{w_{ij}, a_i, b_j\}$ 是 RBM 的参数; w_{ij} 是隐藏层单元与可见层单元之间的连接权重; a_i 和 b_j 分别表示可见层和隐藏层单元的偏置。根据能量函数,可以得出这组状态 (v, h) 的联合概率分布为:

$$P_{\theta}(v, h) = \frac{e^{-E_{\theta}(v, h)}}{Z_{\theta}} \quad (2)$$

$$Z_{\theta} = \sum_{v, h} e^{-E_{\theta}(v, h)} \quad (3)$$

式中: Z_{θ} 为归一化因子。在给定可见层单元 v 或隐藏层 h 的基础上,计算出 RBM 隐藏层单元 h 和可见层单元 v 的条件概率分布分别为:

$$P_{\theta}(h_j = 1 | v) = \text{sigmoid}(b_j + \sum_i v_i w_{ij}) \quad (4)$$

$$P_{\theta}(v_i = 1 | h) = \text{sigmoid}(a_i + \sum_j h_j w_{ij}) \quad (5)$$

式中:

$$\text{sigmoid}(x) = \frac{1}{1 + \exp(-x)} \quad (6)$$

当给定一组训练样本集合 $S = \{v^1, v^2, \dots, v^n\}$ 时,其目标就是最大化如下似然函数:

$$L_{\theta,S} = \sum_{n=1}^N \log P(v^n) \quad (7)$$

式中: N 为训练样本数量。

2.2 高斯-伯努利受限玻尔兹曼机

高斯-伯努利受限玻尔兹曼机 (GBRBM)^[13] 引入高斯函数,使得输入向量不再局限于必须符合伯努利分布(二值分布),解决了 RBM 处理连续输入向量时信息丢失问题。类比 RBM,GBRBM 的能量函数改为:

$$E_{\theta}(v,h) = - \sum_{i=1}^n \frac{(v_i - a_i)^2}{2\sigma_i^2} - \sum_{j=1}^m b_j h_j - \sum_{i=1}^n \sum_{j=1}^m \frac{v_i w_{ij} h_j}{\sigma_i} \quad (8)$$

式中: σ 是高斯函数的标准差; v_i 表示可见层的第 i 个实值输入向量,即第 i 个油耗影响因素输入向量; h_j 仍为二值,即 $h_j \in \{0,1\}$; m 为可见层单元数目,即油耗影响因素输入向量个数; n 为隐藏层的单元数目。根据能量函数,得到 GBRBM 可见层单元和隐藏层单元的条件概率分别为:

$$P_{\theta}(v_i = v | h) = N(a_i + \sigma_i \sum_j w_{ij} h_j, \sigma_i^2) \quad (9)$$

$$P_{\theta}(h_j = 1 | v) = \text{sigmoid}(b_j + \sum_i \frac{v_i w_{ij}}{\sigma_i}) \quad (10)$$

文中 $N(\mu, \sigma_i)$ 为具有均值 μ 和标准差 σ 的高斯函数。GBRBM 的训练过程与 RBM 一样,均采用无监督训练机制进行训练。

2.3 GBRBM-DBN 油耗估计模型架构

本文提出的 GBRBM-DBN 油耗估计模型,根据潘广源等^[14] 提出的基于重构误差的 DBN 网络深度确定方法,确定隐藏层层数为 3。因此 GBRBM-DBN 油耗估计模型包含一个 GBRBM、两个 RBM,并在顶层连接一个 BP 神经网络,GBRBM 的可见层接收油耗影响因素输入向量,其隐藏层作为第一个 RBM 的可见层。同样,第一个 RBM 的隐藏层作为第二个 RBM 的可见层,经过逐层非线性变换后,再通过 BP 网络进行油耗估计。建立的 GBRBM-DBN 油耗估计模型架构如图 2 所示。

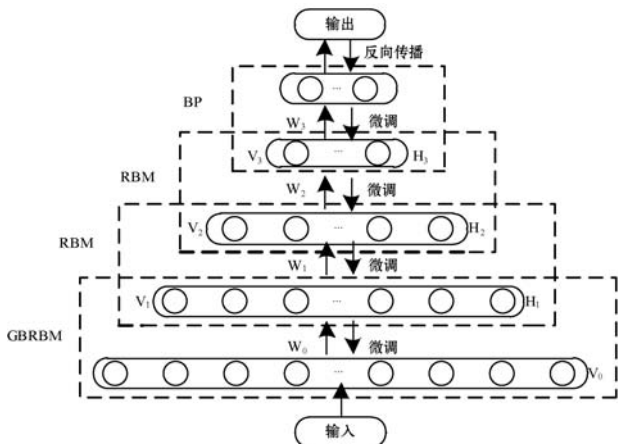


图 2 GBRBM-DBN 油耗估计模型架构

2.4 CD-K 训练算法中的自适应步长

GBRBM-DBN 油耗估计模型利用 K 步对比散度 (CD-K) 算法分别对 GBRBM 和 RBM 进行预训练^[15], 参数 θ 更新公式如下:

$$\theta_t = \theta_{t-1} + \eta \frac{\partial \ln L_{\theta,S}}{\partial \theta} \quad (11)$$

$$\frac{\partial \ln L_{\theta,S}}{\partial \theta} = - \sum_h P_{\theta}(h | v) \frac{\partial E_{\theta}(v,h)}{\partial \theta} + \sum_{v,h} P_{\theta}(v,h) \frac{\partial E_{\theta}(v,h)}{\partial \theta} = \langle \frac{\partial E_{\theta}(v,h)}{\partial \theta} \rangle_{P_{\theta}(h|v)} - \langle \frac{\partial E_{\theta}(v,h)}{\partial \theta} \rangle_{P_{\theta}(v,h)} \quad (12)$$

式中: η 表示步长; $\langle \cdot \rangle_P$ 表示概率分布 P 的期望。但在实际情况下联合概率分布 $P_{\theta}(v,h)$ 的期望很难求出,因此利用 CD 算法中的 K 步 Gibbs 采样方法得到其近似值。

由于在 CD 算法中每个 RBM 均需要进行多次迭代,且每次迭代后的参数更新方向不尽相同,所以需要选择适当的步长 η 。如果选择的 η 较大,容易引起振荡;如果 η 较小,则会导致收敛速度缓慢。为解决这个问题,提出了一种自适应步长 (AS) 的方法^[16]。

AS 方法是根据 RBM 训练过程中连续两次迭代后的参数更新方向来调整 η ,将 $\frac{\partial \ln L_{\theta,S}}{\partial \theta}$ 简记为 $\Delta\theta$,即:

$$\eta = \begin{cases} \alpha\eta' & \Delta\theta_{t-1} \cdot \Delta\theta_{t-2} > 0 \\ \beta\eta' & \Delta\theta_{t-1} \cdot \Delta\theta_{t-2} < 0 \end{cases} \quad (13)$$

式中: $\alpha > 1$ 表示 η 的增加系数; $\beta < 1$ 表示 η 的减小系数,且 $0 < \beta < \alpha$ 。如果两次连续的参数更新方向相同,则 η 增加;如果方向相反,则 η 减小,从而避免了步长选择不当造成的问题,并提高了 CD-K 算法的收敛速度。

2.5 GBRBM-DBN 油耗估计算法

本文利用 GBRBM-DBN 模型进行油耗估计,模型输入是下降段油耗影响因素,设第 i 个油耗影响因素输入向量为 $X_i = [x_{i1}, x_{i2}, \dots, x_{ij}]$,其中 x_{ij} 为第 i 个影响因素的第 j 个输入样本值,模型输出是下降段油耗估计值 Y 。对输入数据进行归一化转化为相同量纲,归一化公式如下:

$$\hat{x}_{ij} = \frac{x_{ij} - (x_{ij})_{\min}}{(x_{ij})_{\max} - (x_{ij})_{\min}} \quad (14)$$

式中: \hat{x}_{ij} 为输入样本归一化后的值; $x_{i\max}$ 和 $x_{i\min}$ 分别为第 i 个油耗影响因素输入样本向量元素的最大值和最小值。

在 GBRBM-DBN 模型中,将归一化的油耗数据赋值给 GBRBM 的可见层单元,通过 GBRBM 和 RBM 对输入向量 X 进行非线性变换,重构到另外一个向量空

间 Z , 即 $H(X) \rightarrow Z$ 。然后给定一组向量空间 Z 与目标油耗 Y 的配对组合 $\{(Z_1, Y_1), (Z_2, Y_2), \dots, (Z_n, Y_n)\}$ 来学习 BP 网络模型 $Y_{n+1} = G(Z)$, 通过最小化目标损失函数 L , 得到估计模型的适合参数 W :

$$L_w = \|Y - G(Z)\|^2 \quad (15)$$

GBRBM-DBN 油耗估计模型的训练分为预训练和反向微调两部分, 过程如下:

(1) 预训练阶段。用 CD-K 算法分别对 GBRBM 和 RBM 进行无监督预训练, K 一般取值为 1; 在预训练过程中, 采用自适应步长 (AS) 方法更新步长 η ; 更新参数 θ , 这里 $\theta = \{w_{ij}, a_i, b_j\}$ 。

(2) 反向微调阶段。在模型顶层利用 BP 算法进行自上而下的参数微调^[17]。由于模型在预训练阶段已经得到模型参数, 在利用 BP 算法参数微调时只需要对已知参数空间进行局部搜索即可^[18]。具体的 GBRBM-DBN 油耗估计算法流程如图 3 所示。

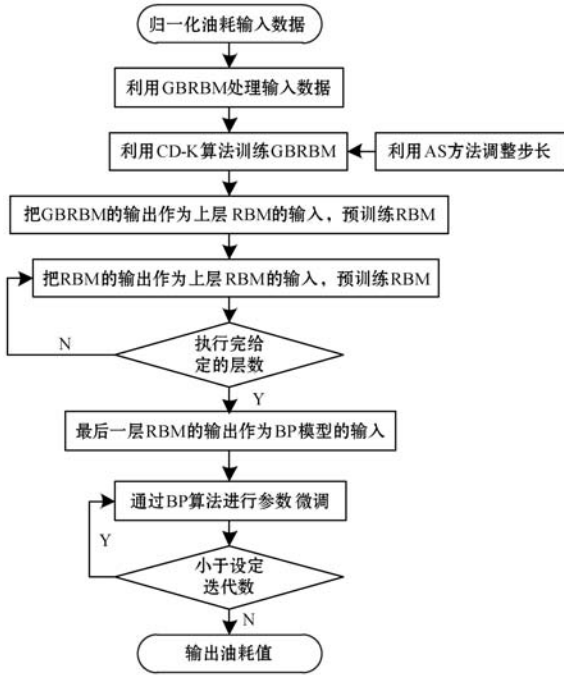


图3 算法流程

在 GBRBM-DBN 油耗估计模型训练完成后, 给定一组油耗影响因素输入数据, 便可以得到对应的油耗估计值。

3 实验仿真

3.1 实验设置

本文实验数据来源于某航空公司同一航线和机型的下降段 QAR 数据, 选取飞机总重、下降距离、下降率、风速、风向、大气总温、俯仰角、迎角、左发转速、右发转速等 10 个参数作为油耗估计模型输入, 以油耗值

作为输出。总共获得 8 000 多组下降段数据, 以 80% 数据作为训练样本, 20% 数据作为测试样本。

文中选取 10 个油耗影响因素作为模型输入, 因此 GBRBM 的可见层节点数为 10, 隐藏层节点数为 8, 两个 RBM 的隐藏层节点数逐层递减, 分别为 6 和 4, BP 神经网络的节点数为 3。经过多次实验发现, 当迭代次数大于 2 000 后, 均方误差的值不再发生变化, 因此设定迭代次数为 2 000。

3.2 性能指标

对于 GBRBM-DBN 油耗估计模型, 本文以均方根误差 (RMSE)、平均绝对百分比误差 (MAPE) 以及拟合优度值 (R) 作为评价体系, 这三种指标的定义分别为:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (16)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (17)$$

$$R = \frac{\sum_{i=1}^N (y_i - \bar{y}_i)^2}{\sum_{i=1}^N (\hat{y}_i - \bar{y}_i)^2} \quad (18)$$

式中: y_i 为下降段油耗估计值; \hat{y}_i 为下降段油耗实际值; \bar{y}_i 为下降段油耗估计平均值。

3.3 实验结果及比较分析

3.3.1 实验结果

实验程序编码在 MATLAB 2009a 平台上完成, 计算机处理器为 Intel-I7, 内存为 8 GB。将选取出的油耗影响因素实验数据 X 归一化后输入到 GBRBM-DBN 油耗估计模型中, 通过 GBRBM 和 RBM 进行非线性变换得到 $Z = H(X)$, 再利用顶层连接的 BP 神经网络对变换后的数据进行油耗估计得到油耗估计值 $Y = G(Z)$ 。仿真实验结果如图 4 和图 5 所示。

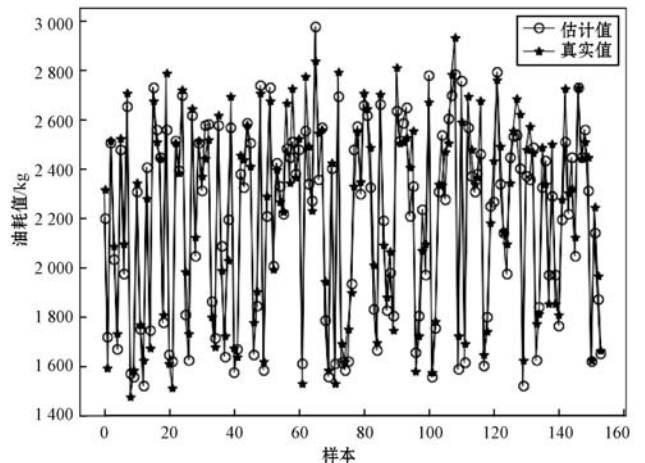


图4 油耗估计值和实际值

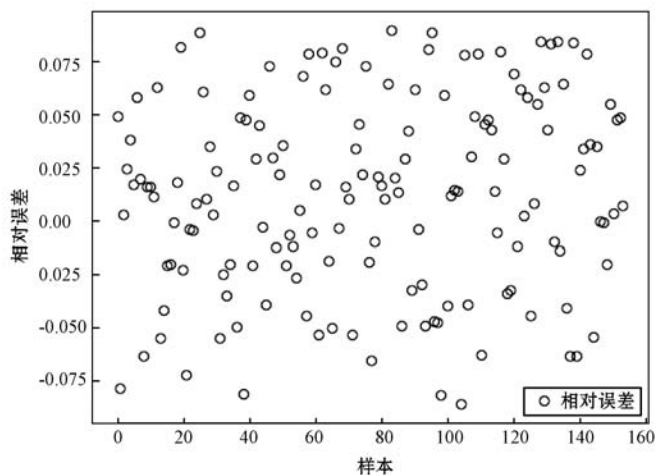


图 5 相对误差

图 4 为部分样本的油耗估计值和实际值的对比结果,图 5 为油耗估计值和实际值的相对误差。GBRBM-DBN 模型各评价指标值如下:均方根误差为 58.74,平均绝对百分比误差为 2.67%,拟合优度为 0.972 3。

根据图 4 和图 5 以及评价指标值可以看出,GBRBM-DBN 估计模型的油耗估计值与实际值偏差较小,考虑到下降段油耗还会受到天气状况和航路拥堵等其他因素的影响,所以允许误差存在,本文提出的油耗估计模型在误差允许范围内具有可行性。

3.3.2 不同步长时间复杂度比较

在不同步长时间复杂度的对比实验中,用自适应步长与三种不同的固定步长($\eta = 0.1, \eta = 0.5, \eta = 0.9$)进行比较。对于自适应步长来说,初始步长设置为 0.1,增加系数 α 和减小系数 β 的值分别为 1.2 和 0.6。表 1 为自适应步长与三种固定步长的时间复杂度对比。可以看出,自适应步长优于固定步长。

表 1 不同步长时间复杂度对比

步长	时间复杂度/s
自适应步长	16.27
$\eta = 0.1$	21.51
$\eta = 0.5$	20.35
$\eta = 0.9$	23.74

3.3.3 不同模型油耗估计结果比较

文中分别对 BP 神经网络模型、RBM-DBN 模型以及本文提出的 GBRBM-DBN 模型的油耗估计结果进行了比较。由式(16) - 式(18)分别计算得出这三种模型油耗估计结果的 MSE、MAPE 和拟合优度 R ,计算结果如表 2 所示。

表 2 不同模型结果对比

估计模型	RMSE	MAPE	R
GBRBM-DBN	58.74	2.67%	0.972 3
RBM-DBN	82.53	3.74%	0.953 5
BP	163.17	7.52%	0.910 8

由表 2 可以看出,GBRBM-DBN 模型的非线性拟合能力要优于 BP 神经网络模型和 RBM-DBN 模型,估计精度也有小幅度提高;GBRBM-DBN 模型在飞机下降段油耗估计方面有较大优势。

4 结 语

为准确估计飞机燃油消耗量,本文针对飞机下降阶段提出了一种基于改进 DBN 模型的油耗估计方法。通过实验验证了在引入 GBRBM 后,改进 DBN 油耗估计模型的非线性拟合能力和估计精度均有小幅度提高;在 CD-K 训练算法中加入自适应步长(AS)加快了收敛速度,降低了时间复杂度。分析改进 DBN 油耗估计模型的评价指标值,考虑实际误差存在的原因,可以得出,本文提出的方法是一种有效的飞机下降段油耗估计方法,可为其他飞行阶段的油耗估计提供一定的参考。

参 考 文 献

- [1] 陈静杰,肖冠平. 飞机航段油耗区间估计最小样本量分析[J]. 计算机工程与设计, 2014, 35(12): 4356 - 4359.
- [2] Grote M, Williams I, Preston J. Direct carbon dioxide emissions from civil aircraft [J]. Atmospheric Environment, 2014, 95(1): 214 - 224.
- [3] Tang L, Yu L, Wang S H, et al. A novel hybrid ensemble learning paradigm for nuclear energy consumption forecasting [J]. Applied Energy, 2012, 93(5): 432 - 443.
- [4] 田虎森,谢寿生,任立通,等. 军用飞机燃油消耗多元线性回归模型[J]. 火力与指挥控制, 2014, 39(10): 104 - 107.
- [5] Chang R C. Examination of excessive fuel consumption for transport jet aircraft based on fuzzy-logic models of flight data [J]. Fuzzy Sets and Systems, 2015, 269: 115 - 134.
- [6] 郝云霄,闫楚良,刘克格. 基于支持向量机的机翼载荷模型研究[J]. 科学技术与工程, 2013, 13(25): 7432 - 7437.
- [7] 陈静杰,颜艳,刘家学. 基于 BP 网络巡航段油耗估计及影响结构分析[J]. 计算机测量与控制, 2015, 23(6): 2035 - 2138.
- [8] 陈静杰,张永平. 基于 AGA-BP 强估计的爬升段油耗估计[J]. 计算机工程与设计, 2017, 38(10): 2845 - 2849.

- [9] 张鹏程, 张雷, 王继民. 一种基于深度网络的多环境因素降水量预报模型[J]. 计算机应用与软件, 2017, 34(9): 240-245, 261.
- [10] 李飞, 高晓光, 万开方. 基于动态 Gibbs 采样的 RBM 训练算法研究[J]. 自动化学报, 2016, 42(6): 931-942.
- [11] 张翔, 石力, 尚勃, 等. 深度置信网络的 Spark 并行化在微博情感分类中的应用研究[J]. 计算机应用与软件, 2018, 35(2): 48-53.
- [12] 刘建伟, 刘媛, 罗雄麟. 深度学习研究进展[J]. 计算机应用研究, 2014, 31(7): 1921-1930.
- [13] Karakida R, Okada M, Amari S I. Dynamical analysis of contrastive divergence learning: Restricted Boltzmann machines with Gaussian visible units[J]. Neural Networks, 2016, 79: 78-87.
- [14] 潘广源, 柴伟, 乔俊飞. DBN 网络的深度确定方法[J]. 控制与决策, 2015, 30(2): 256-260.
- [15] Hinton G, Osindero S, Teh Y. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7): 1527-1554.
- [16] Lopes N, Ribeiro B. Towards adaptive learning with improved convergence of deep belief networks on graphics processing units[J]. Pattern Recognition, 2014, 47(1): 114-127.
- [17] Hinton G, Salakhutdinov R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504-507.
- [18] Lü Y, Duan Y, Kang W, et al. Traffic Flow Prediction With Big Data: A Deep Learning Approach[J]. IEEE Transactions on Intelligent Transportation Systems, 2015, 16(2): 865-873.
- [6] Batista G E A P A, Monard M C. An analysis of four missing data treatment methods for supervised learning[J]. Applied Artificial Intelligence, 2003, 17(5/6): 519-533.
- [7] Zhang C, Zhu X, Zhang J, et al. GBKII: An Imputation Method for Missing Values[C]// Proceedings of the 11th Pacific-Asia conference on Advances in knowledge discovery and data mining. Springer-Verlag, 2007: 1080-1087.
- [8] 杨涛, 骆嘉伟, 王艳, 等. 基于马氏距离的缺失值填充算法[J]. 计算机应用, 2005, 25(12): 2868-2871.
- [9] 张红霞. 缺失值填充: 基于信息增益的方法[J]. 计算机工程与设计, 2006, 27(24): 4810-4812.
- [10] 卜范玉, 陈志奎, 张清辰. 基于聚类 and 自动编码机的缺失数据填充算法[J]. 计算机工程与应用, 2015, 51(18): 13-17.
- [11] 杨镇楠. 基于自动编码机的增量聚类算法研究及应用[D]. 大连: 大连理工大学, 2016.
- [12] Wei W, Tang Y. A generic neural network approach for filling missing data in data mining[C]// IEEE International Conference on Systems, Man and Cybernetics. IEEE, 2003: 862-867.
- [13] 王继奎, 李少波. 多数据源冲突的主数据真值发现算法[J]. 计算机工程与设计, 2014, 35(1): 177-182.
- [14] 朱会娟, 蒋同海, 周喜, 等. 基于动态可配置规则的数据清洗方法[J]. 计算机应用, 2017, 37(4): 1014-1020.
- [15] 林海伦, 王元卓, 贾岩涛, 等. 面向网络大数据的知识融合方法综述[J]. 计算机学报, 2017, 40(1): 1-27.
- [16] Ristad E S, Yianilos P N. Learning string-edit distance[J]. IEEE Transactions on Pattern Analysis Machine Intelligence, 1996, 20(5): 522-532.
- [17] Waguih D A, Bertiequille L. Truth Discovery Algorithms: An Experimental Evaluation[EB]. arXiv:1409.6428, 2014.
- [18] Blanco L, Crescenzi V, Merialdo P, et al. Probabilistic Models to Reconcile Complex Data from Inaccurate Data Sources[C]// International Conference on Advanced Information Systems Engineering. Springer-Verlag, 2010: 83-97.
- [19] Wang D, Kaplan L, Le H, et al. On truth discovery in social sensing: a maximum likelihood estimation approach[C]// ACM/IEEE, International Conference on Information Processing in Sensor Networks. IEEE, 2012: 233-244.
- [20] Zhao B, Rubinstein B I P, Gemmel J, et al. A Bayesian approach to discovering truth from conflicting sources for data integration[J]. Proceedings of the VLDB Endowment, 2012, 5(6): 550-561.
- [21] Li Y, Gao J, Meng C, et al. A Survey on Truth Discovery[J]. ACM SIGKDD Explorations Newsletter, 2016, 17(2): 1-16.

(上接第 46 页)

参 考 文 献

- [1] Yin X, Han J, Yu P S. Truth Discovery with Multiple Conflicting Information Providers on the Web[C]// Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2007: 1048-1052.
- [2] 金连. 不完全数据中缺失值填充关键技术研究[D]. 哈尔滨: 哈尔滨工业大学, 2013.
- [3] 杨军, 赵宇, 丁文兴. 抽样调查中缺失数据的插补方法[J]. 数理统计与管理, 2008, 27(5): 821-832.
- [4] 李宏, 阿玛尼, 李平, 等. 基于 EM 和贝叶斯网络的丢失数据填充算法[J]. 计算机工程与应用, 2010, 46(5): 123-125.
- [5] Feng H, Chen G, Cheng Y, et al. A SVM Regression Based Approach to Filling in Missing Values[J]. Lecture Notes in