

基于DBN的不均衡样本驱动民航 发动机故障诊断

钟诗胜^{1,2}, 李旭², 张永健²

(1. 哈尔滨工业大学 机电工程学院, 哈尔滨 150001;

2. 哈尔滨工业大学(威海) 船舶与海洋工程学院, 山东 威海 264200)

摘 要: 在结合深度置信网络(DBN)、采样与集成技术的基础上,提出了基于不均衡样本驱动的民航发动机故障诊断模型。该模型通过分析民航发动机历史飞行数据,利用DBN提取性能参数中的内部特征,利用采样技术将不均衡样本均衡化,采用集成技术进行故障分类。将该模型应用到CFM56-7B系列发动机历史飞行数据,实验结果表明:与常用故障诊断方法相比,该模型的准确率高达0.996,AUC值高达0.948,可以有效处理民航发动机样本高维、不均衡问题。

关 键 词: 民航发动机; 故障诊断; 不均衡样本; 深度置信网络; Adaboost.M1 算法

中图分类号: V263.6

文献标志码: A

Fault diagnosis of civil aero-engine driven by unbalanced samples based on DBN

ZHONG Shisheng^{1,2}, LI Xu², ZHANG Yongjian²

(1. School of Mechatronics Engineering,

Harbin Institute of Technology, Harbin 150001, China;

2. School of Naval Architecture and Ocean Engineering,

Harbin Institute of Technology (Weihai), Weihai Shandong 264200, China)

Abstract: Through combination of deep belief network (DBN), sampling and integration technology, a fault diagnosis model of civil aero-engine based on unbalanced sample driving was proposed. By analyzing the historical flight data of civil aero-engines, the model used DBN to extract the internal features of the performance parameters, then used the sampling technology to equalize the unbalanced samples, and finally adopted integrated technology for fault classification. The model was applied to historical flight data of CFM56-7B series engines. Compared with common fault diagnosis methods, the experimental results showed that the model had higher accuracy of 0.996 and AUC value of 0.948, and can effectively deal with high-dimensional and unbalanced problems of civil aero-engine samples.

Key words: civil aero-engine; fault diagnosis; unbalanced samples;
deep belief network; Adaboost.M1 algorithm

收稿日期:2018-07-09

基金项目:国家自然科学基金重点项目(U1533202); 民航科技项目(MHRD20150104)

作者简介:钟诗胜(1964—),男,教授、博士生导师,博士,主要从事航空发动机健康管理,智能制造技术,数控技术与应用等研究。

引用格式:钟诗胜,李旭,张永健. 基于DBN的不均衡样本驱动民航发动机故障诊断[J]. 航空动力学报,2019,34(3):708-716.

ZHONG Shisheng, LI Xu, ZHANG Yongjian. Fault diagnosis of civil aero-engine driven by unbalanced samples based on DBN[J]. Journal of Aerospace Power, 2019, 34(3): 708-716.

民航发动机作为飞机上最主要的部件之一,对其进行故障诊断研究,不仅可以保证飞机的安全运行,也可以保证生命及财产安全^[1-2]。民航发动机故障主要有气路故障、机械零部件故障以及附件故障三种^[3]。据中国民航科学技术研究院对我国 2011~2015 年民航事故统计,气路故障大约占民航发动机故障的 60%^[4]。近年来,很多学者针对民航发动机气路故障诊断做了深入研究,总结其研究方法,大致可以分为三类:基于模型的故障诊断方法、基于知识的故障诊断方法和基于数据的故障诊断方法。随着机器学习、深度学习等人工智能技术的快速发展,基于数据的故障诊断方法通过构建智能算法模型挖掘发动机数据间的内部特征,越来越广泛地应用到发动机状态监测中^[5-9]。

民航发动机数据参数多、参数之间关系复杂、正常样本与故障样本比例不均衡。传统的发动机故障诊断方法在处理发动机数据时存在两个问题:一方面,传统的特征提取方法如主成分分析(principal component analysis, PCA)、线性判别分析(linear discriminant analysis, LDA)等难以挖掘民航发动机数据间的深层特征;另一方面,传统的分类器大部分以最大化全部样本分类准确率为目标,难以保证比较关注的故障样本有很好的准确率^[10-13]。在工程应用领域,一般认为多数类与少数类样本数量比高于 2:1 时,样本分布具有不均衡性^[14]。目前,针对不均衡问题常用解决方法有一对多法(one-vs-all, OVA)和一对一法(one-vs-one, OVO)^[15]。但这两种方法在解决样本不均衡问题时不仅需要很多的分类器,而且并不能从根本上解决样本不均衡问题。因此,需要一种新技术来挖掘民航发动机数据内部特征并解决类别不均衡问题。

针对上述问题,受深度学习、采样与集成学习技术启发,本文提出一种基于不均衡样本驱动的民航发动机气路故障诊断方法,把收集到的高维发动机数据输入到深度置信网络(deep belief network, DBN)进行深度特征提取;然后,通过下采样技术,将不均样本分割成多个均衡样本子集;最后,利用集成学习框架进行分类。该方法在结合以上三种技术的基础上,构建民航发动机气路故障诊断模型,避免了传统数据驱动模型在解决高维、不均衡问题时的不足。

1 特征提取与分类模型

1.1 DBN 的结构与训练过程

深度学习的本质,就是将原始数据经过多层神经网络的非线性映射转换成更高层次、更抽象的特征数据,从而对原始数据进行挖掘分析。因此,与人工特征提取和浅层学习特征提取相比,深度学习更能发现数据的内在信息,更具代表性。DBN 作为深度学习算法之一,有着良好的特征提取能力与分类能力。

DBN 模型结构如图 1 所示,DBN 模型由多个受限玻耳兹曼机(restricted Boltzmann machine, RBM)与顶层分类器组成^[16],分别对应 DBN 模型训练的两个阶段:RBM 无监督训练和分类器的有监督微调。无监督训练过程通过逐层训练 RBM 实现,得到各层 RBM 的参数;然后通过分类器的有监督微调,将误差反向传播,对模型参数进行修正。图中 v 为可视层向量, h 为隐藏层向量。

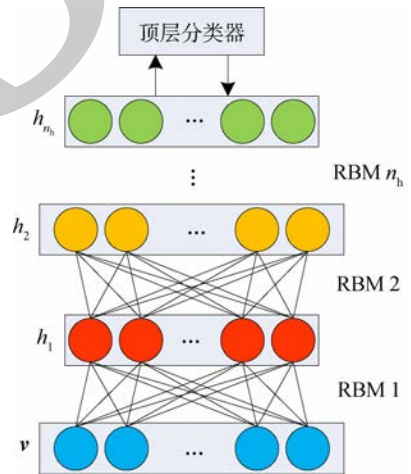


图 1 DBN 结构

Fig. 1 Structure of DBN

1.1.1 RBM 及其训练

RBM 作为 DBN 的基本单元,其网络结构如图 2 所示。RBM 包含可视层和隐藏层,层内节点之间没有连接,相互独立,层间节点之间全连接。输入向量可以通过可视层 v 输入到 RBM 网络,隐藏层 h 输出的向量可以作为原始向量的另一种表示,从而实现特征提取。

可视层向量: $v = (v_1, v_2, \dots, v_{n_v})^T \in \mathbf{R}^{n_v}$

隐藏层向量: $h = (h_1, h_2, \dots, h_{n_h})^T \in \mathbf{R}^{n_h}$

可视层偏置: $a = (a_1, a_2, \dots, a_{n_v})^T \in \mathbf{R}^{n_v}$

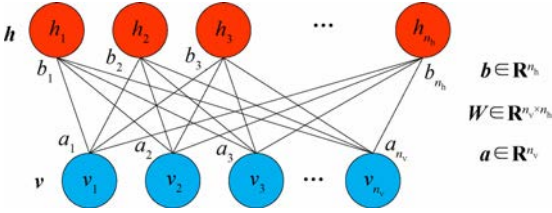


图2 RBM结构

Fig. 2 Structure of RBM

隐藏层偏置: $\mathbf{b} = (b_1, b_2, \dots, b_{n_h})^T \in \mathbf{R}^{n_h}$

连接权重: $\mathbf{W} = (w_{i,j}) \in \mathbf{R}^{n_v \times n_h}$

其中 n_v, n_h 分别表示可视层与隐藏层神经元的数目; a_i, b_j 分别表示可视层第 i 个神经元与隐藏层第 j 个神经元的偏置; $w_{i,j}$ 表示可视层第 i 个神经元与隐藏层第 j 个神经元之间的连接权值。

RBM 中每个神经元含有两种状态: 激活(用 1 表示)、未激活(用 0 表示)。根据 RBM 结构, 对于任意给定的一组状态 $(\mathbf{v}, \mathbf{h})$, 可以定义其能量函数为

$$E(\mathbf{v}, \mathbf{h}) = - \sum_{i=1}^{n_v} a_i v_i - \sum_{j=1}^{n_h} b_j h_j - \sum_{i=1}^{n_v} \sum_{j=1}^{n_h} v_i w_{i,j} h_j \quad (1)$$

在给定可视层状态的基础上, 根据可视层与隐藏层的联合概率分布和可视层与隐藏层内各节点条件独立, 可以求得隐藏层节点 h_j 与可视层节点 v_i 的激活概率为

$$p(h_j = 1 | \mathbf{v}) = \text{sigmoid} \left[b_j + \sum_{i=1}^{n_v} (v_i w_{i,j}) \right] \quad (2)$$

$$p(v_i = 1 | \mathbf{h}) = \text{sigmoid} \left[a_i + \sum_{j=1}^{n_h} (w_{i,j} h_j) \right] \quad (3)$$

RBM 模型通过极大似然方法进行学习, 利用交替吉布斯采样方法求解参数 $\theta = \{\mathbf{a}, \mathbf{b}, \mathbf{W}\}$, 让 RBM 在参数 θ 作用下表示的概率分布尽可能地拟合输入样本的概率分布。然而该方法的收敛速度与采样次数难以确定, 导致 RBM 训练时间非常长。因此, Hinton 教授提出了对比散度 (contrastive divergence, CD) 方法, 又称为 CD- k 算法^[17], 通常采样次数 $k=1$ 就可以得到比较好的估计结果。利用 CD- k 算法进行参数更新, 更新公式为

$$\Delta \mathbf{W} = \Delta \mathbf{W} +$$

$$[\rho(\mathbf{h} = \mathbf{1} | \mathbf{v}^{(0)}) \mathbf{v}^{(0)} - \rho(\mathbf{h} = \mathbf{1} | \mathbf{v}^{(k)}) \mathbf{v}^{(k)}]$$

$$\Delta \mathbf{a} = \Delta \mathbf{a} + (\mathbf{v}^{(0)} - \mathbf{v}^{(k)})$$

$$\Delta \mathbf{b} = \Delta \mathbf{b} +$$

$$[\rho(\mathbf{h} = \mathbf{1} | \mathbf{v}^{(0)}) \mathbf{v}^{(0)} - \rho(\mathbf{h} = \mathbf{1} | \mathbf{v}^{(k)}) \mathbf{v}^{(k)}] \quad (4)$$

1.1.2 DBN 模型的训练

在训练 DBN 模型时, 上一层 RBM 的输出作为下一层 RBM 的输入, 在此过程中, 上一层 RBM 的误差会传递到下一层。因此, 需要采用有监督学习对误差进行反向传播, 调节参数。本文采用比较常用的 softmax 分类器调节 DBN 模型的参数, DBN 整体训练过程如下:

步骤 1 参数初始化。包括权重矩阵 $\mathbf{W}$, 可视层偏置向量 $\mathbf{a}$, 隐藏层偏置向量 $\mathbf{b}$, 采样次数 k , 迭代次数 J , 学习率 l_r 的初始化。

步骤 2 参数更新。利用 CD- k 算法, 进行 k 次吉布斯采样, 利用参数更新公式更新参数。

步骤 3 逐层训练。逐层训练每个 RBM, 直至所有 RBM 训练完成。

步骤 4 微调。利用分类器的误差反向传播机制, 调整各层网络的权值与偏置。

1.2 下采样技术

采样技术包括下采样技术与过采样技术两种, 两者均可将不平衡样本均衡化。但是, 过采样技术通过多次复制少数类样本来增加少数样本数量, 在模型训练过程中容易出现过拟合现象, 因此使用较少^[18]。下采样技术的核心思想是从多数类数据集中抽取与少数类样本数目一致的样本, 并与少数类样本组合成新的样本子集, 从而使不平衡样本均衡化。为保证采样得到的每个样本的随机性与独立性, 选取有放回采样, 实现过程如下:

步骤 1 对于输入的不均衡的训练数据集 $S = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$, $x_i \in \mathbf{R}$, $y_i \in C$, 类别 $C = \{c_1, \dots, c_{N_c}\}$, 选取采样次数 T , 从而得到均衡的样本子集。

步骤 2 从多数类样本集中进行有放回采样并与少数类样本合并得到 T 个子集 $S = \{S_1, S_2, \dots, S_T\}$, 对 $\forall S_i \in S$, S_i 为均衡样本集。

1.3 集成模型

集成学习可以融合多个分类模型, 提高模型的鲁棒性, 而且与采样算法可以很好的结合。Adaboost.M1 为 Adaboost 算法中的一种, 可以很好地解决多类别标签问题。Adaboost.M1 算法通过调节每个样本的权重进行学习^[19-20]。如图 3 所示, 实现该算法首先需要初始化样本权重, 然后改变样本的权重来凸显样本作用的大小。对于分类错误的样本, 增大其对应的权重, 反之, 减小

其对应的权重系数。图中 $W_{1,i}, \dots, W_{M,i}$ 表示样本权重系数; $G_1, \dots, G_M$ 表示分类器; $\alpha_1, \dots, \alpha_M$ 表示分类器的权重系数。

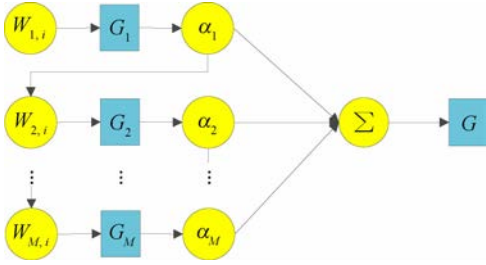


图 3 Adaboost, M1 算法原理图

Fig. 3 Schematic diagram of Adaboost, M1 algorithm

由于 Adaboost, M1 模型集成了多个单一学习模型,所以相比较单一学习模型而言,集成模型在实际应用中具有良好的表现能力,具有更高的鲁棒性。本文使用 Adaboost, M1 集成模型将多个支持向量机(SVM)弱学习器进行组合,构造一个强分类器。通过调节弱学习器权值的大小,改变其在表决过程中作用的大小。对于输入的数据 $S = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$, $x_i \in \mathbf{R}$, $y_i \in C$, $C = \{c_1, \dots, c_{N_c}\}$, Adaboost, M1 实现流程如下:

步骤 1 初始化训练数据的权值分布。

$$\beta_{1,i} = \frac{1}{N}, \quad i = 1, 2, \dots, N \quad (5)$$

步骤 2 对于 $m = 1, 2, \dots, M$, 通过计算 G_m 的分类误差率 e_m 更新 G_m 的权重系数 α_m , 然后通过 α_m 更新样本权值 $\beta_{m+1,i}$, 直至更新完所有分类器, 其中 $i = 1, 2, \dots, N$ 。

$$e_m = \sum_{i=1}^N \beta_{m,i} I(G_m(x_i) \neq y_i) \quad (6)$$

$$\alpha_m = \frac{1 - e_m}{e_m} \quad (7)$$

$$\beta_{m+1,i} = \frac{\beta_{m,i}}{Z_m} \exp[-\alpha_m y_i G_m(x_i)] \quad (8)$$

式中 $Z_m = \sum_{i=1}^N \beta_{m,i} \exp[-\alpha_m y_i G_m(x_i)]$ 表示归一化因子。

步骤 3 通过对各个弱分类器线性组合, 构建强分类器。

$$G(x) = \arg \max_{y \in C} \sum_{m=1}^M \ln\left(\frac{1}{\alpha_m}\right) \quad (G_m = y) \quad (9)$$

1.4 模型评价指标的选取

1.4.1 DBN 特征提取模型评价指标

特征提取作为故障诊断的前提, 提取特征的

好坏直接关系到故障诊断精度的高低。目前常用于特征提取能力的评价指标有重构误差、失真度、时间等。本文选取重构误差和训练时间作为评价指标, 重构误差 e_{rms} 计算公式为

$$e_{rms} = \sqrt{\frac{\sum_{i=1}^N \sum_{j=1}^{N_d} (x_i^j - y_i^j)^2}{NN_d}} \quad (10)$$

式中 N 表示样本个数, N_d 表示每个样本的特征维数。

1.4.2 集成分类模型评价指标

由于民航发动机正常样本与故障样本数目往往是不均衡的, 传统故障诊断算法默认样本是均衡的, 忽略了小类别样本更加重要的特性。直接采用不均衡样本对模型进行训练可能导致高准确率模型很差的故障诊断能力, 难以应用到工程应用中。例如正负样本比例为 9 : 1 时, 当分类器把负类样本全部诊断为正类样本, 虽然可以达到 90% 的正确率, 但是这样的分类器往往没有价值, 因为它不能预测出负类样本。所以为了更好地评价分类器对不均衡样本的分类能力, 在准确率的基础上, 同时采用混淆矩阵与 ROC(receiver operating characteristic curve) 进行评价。ROC 反映了真阳率与假阳率的关系, 曲线下的面积 (area under curve, AUC, 记为 A_{uc}) 给出了分类器的综合性能, 一个完美分类器的 AUC 为 1.0, 随机猜测的 AUC 为 0.5。所以, AUC 的值越大, 分类器的性能越好。准确率、真阳率、假阳率的计算公式为

$$A_{cc} = \frac{T_p + T_n}{T_p + F_p + T_n + F_n} \quad (11)$$

$$T_{pr} = \frac{T_p}{T_p + F_n} \quad (12)$$

$$F_{pr} = \frac{F_p}{T_n + F_p} \quad (13)$$

式中 T_p 为准确识别出的故障样本的数量; F_p 为将正常样本识别为故障样本的数量; F_n 为将故障样本识别为正常样本的数量; T_n 为准确识别出正常样本的数量。

2 基于 DBN 的民航发动机故障诊断流程

基于不均衡样本驱动的民航发动机故障诊断模型主要包括 DBN 特征提取模型与 Adaboost, M1 集成分类器两部分, 先利用训练集对 DBN 与集成模型进行训练, 然后把训练好的故障诊断模

型应用到测试集进行评估。本文模型框架如图 4 所示。

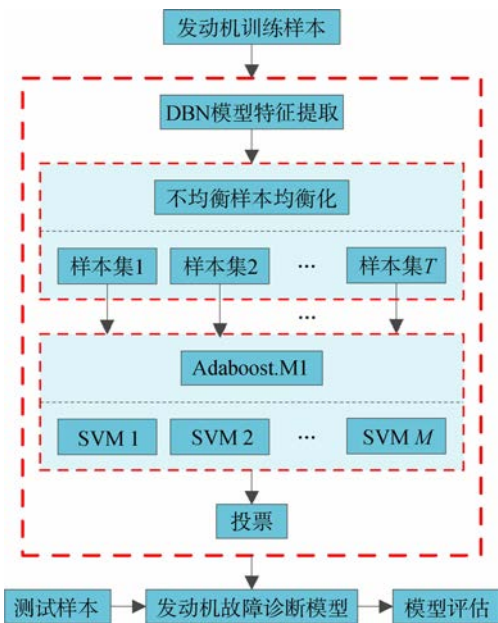


图 4 本文模型故障诊断框架

Fig. 4 Fault diagnosis framework of the proposed model

基于 DBN 的不均衡样本驱动民航发动机故障诊断方法详细步骤如下：

- 步骤 1 收集故障样本数据，构建故障样本集，并进行训练集与测试集的划分。
- 步骤 2 利用训练集对 DBN 特征提取模型训练，选取 DBN 结构参数。
- 步骤 3 对不均衡样本进行 T 次采样，得到 T 个样本子集，并对每个样本子集构建一个包含 M 个 SVM 的集成分类器。
- 步骤 4 采用投票决策的方式，票数最多的类别作为最终类别。
- 步骤 5 利用训练好的 DBN 特征提取模型与集成分类器对测试集进行故障诊断。
- 步骤 6 统计模型故障诊断结果，选取相应指标进行评估。

3 实验验证

3.1 民航发动机故障样本的收集

本文以 CFM56-7B 系列发动机机队为样本机队，通过分析 CNR(customer notification report)发现，OEM(original equipment manufacturer)厂家通常利用排气温度变化量(delta exhaust gas temperature,DEGT)、排气温度裕度变

化(exhaust gas temperature margin,EGTM)、核心机转速变化量(delta core speed,DN2)、燃油流量变化值(delta fuel flow,DFF)4 种性能参数结合指印图进行诊断^[21]。因此，本文选取 DEGT、EGTM、DFF 和 DN2 作为气路故障指征参数。根据 OEM 厂家对故障预报数据分析，其选取的故障征候循环数从 5 到 130 不等，且只有少数故障征候循环数超过 10。因此，实验选取 10 个飞行循环区间内的数据作为一个样本。通过整理发动机历史飞行数据，收集到 1 661 组发动机样本，结合本文选取的 DEGT、EGTM、DFF 和 DN2 共 4 个故障指征参数，因此每个样本的维数为 40 维。采集的样本如表 1 所示。

表 1 发动样本数据
Table 1 Engine sample set

组号	排气 温度 变化量	核心机 转速 变化量	燃油 流量 变化值	排气温度 裕度 变化量	状态
33.6549	−0.4335	−0.7509	33.3924		正常
33.1272	−0.4606	−0.4783	38.5706		
32.0203	−0.0162	0.4767	33.6118		
33.1211	−0.6518	−0.5428	31.4644		
33.2989	−0.5609	0.7513	16.9258		
33.5979	−0.4546	−0.5839	29.2183		
32.6797	−0.5601	−0.5917	32.1233		
34.9597	−0.7956	0.4235	26.8326		
35.2779	−0.5154	0.6485	27.9856		
35.5231	−0.4939	0.4315	32.5973		
⋮	⋮	⋮	⋮	⋮	
33.354	−0.5017	−0.7998	42.42		低压 涡轮 烧蚀
31.7788	−0.3496	−1.1295	29.2217		
34.3956	−0.4535	−0.7953	31.4429		
34.928	−0.2639	0.7203	34.8494		
31.0808	−0.7168	−0.3538	31.0054		
38.0394	−0.2898	−0.4124	36.7673		
34.3068	−0.6333	−0.581	26.3474		
34.7128	−0.5057	−0.4868	30.897		
31.8918	−0.3879	−0.8021	28.3649		
35.8884	−0.3696	1.0422	36.8509		

3.2 建立 DBN 特征提取模型

为方便数据处理与消除参数间量纲的影响，

本文将原始数据进行归一化。首先根据归一化后样本维数和样本类别数确定 DBN 输入层为 40 个节点,输出层为 3 个节点,然后构建 3 层、4 层、5 层的 7 个 DBN 网络结构,然后通过重构误差和时间指标选取最优 DBN 模型。

本部分实验选取 900 组正常样本对模型进行无监督训练,通过实验,可以得到各个网络结构的重构误差与训练时间如表 2 所示,可以发现,当网络结构为[40,30,3]时重构误差最小,训练时间与用时最短的[40,10,3]相差不大,可以认为相对于其余 6 种网络结构,[40,30,3]最优,即该网络结构下 DBN 的特征提取能力最好。DBN 模型参数设置如下:迭代次数为 500 次,学习率为 0.001, k 取 1。网络结构为[40,30,3]时,原始曲线与重构曲线如图 5 所示。

表 2 DBN 结构与实验结果

Table 2 DBN structure and experimental results

DBN 结构	重构误差	训练时间/s
[40,10,3]	0.11	6.133
[40,20,3]	0.106	8.419
[40,30,3]	0.105	10.788
[40,20,10,3]	0.126	13.422
[40,30,10,3]	0.121	15.317
[40,30,20,3]	0.117	18.273
[40,30,20,10,3]	0.143	22.633

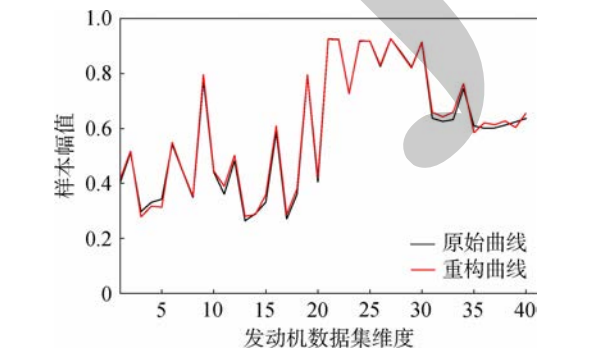


图 5 DBN 结构为[40,30,3]时重构曲线与原始曲线

Fig.5 Reconstruction curve and original curve with DBN structure of [40, 30, 3]

3.3 建立集成分类模型

由于民航发动机样本中正常样本占大多数,故障样本很少,在类别不均衡的样本中,少数类更重要。然而,传统数据驱动的方法隐含假设每个

样本的重要性是相同的,所以并不适用于不均衡样本。因此,需要新的方法来进行不均衡样本的故障识别。本部分实验共选取 761 组样本,正常样本与故障样本比例远大于 2 : 1,属于不均衡样本,训练集与测试集划分如表 3 所示。

表 3 发动机样本分配情况

Table 3 Distribution of the sample engine

样本	训练集	测试集
正常样本	490	210
高压涡轮烧蚀	21	9
低压涡轮烧蚀	21	10

为了验证提出模型针对不均衡样本的分类能力,为本文模型设置 3 组对比实验,分别是:DBN、SVM、DBN+SVM,模型参数设置参看表 4。为消除算法的随机性,每个模型进行 10 次实验,结果取平均值,统计 4 个模型对不均衡数据的分类能力,结果如表 5 所示。通过表 5 可以发现 4 个分类模型的准确率均很高,差别不是很明显。但是,通过对比各个模型的 AUC 值,本文提出模型的 AUC 值为 0.948,大于其余三类模型的 AUC 值,说明本文提出的模型在不均衡样本分类过程中有较好的分类能力。

表 4 模型及参数

Table 4 Models and parameters

模型	参数
DBN	$k=1, l_{r1}=0.001, J=500$
SVM	$\gamma=1.0, l_{r2}=0.0001, J=500$
DBN+SVM	$k=1, l_{r1}=0.001, J=500,$ $\gamma=1.0, l_{r2}=0.0001, J=500$
本文模型	$k=1, l_{r1}=0.001, J=500,$ $T=10, \text{弱分类器选 SVM}, M=5,$ $\gamma=1.0, l_{r2}=0.0001, J=500$

注: γ 为 SVM 的惩罚因子。

表 5 诊断结果对比

Table 5 Comparison of diagnostic results

模型	准确率	A_{uc}
DBN	0.917	0.500
SVM	0.965	0.633
DBN+SVM	0.961	0.927
本文模型	0.996	0.948

为了更加直观地展示各个模型的实验效果,各个模型的混淆矩阵如图 6 所示、各个模型的 ROC 如图 7~图 10 所示。观察图 6 中 DBN 与 SVM 模型的混淆矩阵与图 7、图 8 的平均 ROC,可以发现 DBN 分类器对于不平衡样本没有分类能力,而 SVM 有一定的分类能力;观察图 6 中 SVM 与 DBN+SVM 模型的混淆矩阵与图 8、图 9 的平均 ROC,可以发现,SVM 利用经过 DBN 提取的特征作为输入比单一的 SVM 利用原始数据作为输入分类能力强。观察图 6 中 DBN+SVM 与本文模型的混淆矩阵与图 9、图 10 的平均 ROC,可以发现,本文模型的分类能力要好于 DBN+SVM,说明将各个子集分类器集成后,分类能力强于单一子集分类器。通过上述对比实验,说明了本文模型在结合 DBN、采样与集成技术之后,可以有效利用原始数据的高维特征和处理样本的不均衡性,对于高维、不平衡样本有很好的分类能力。

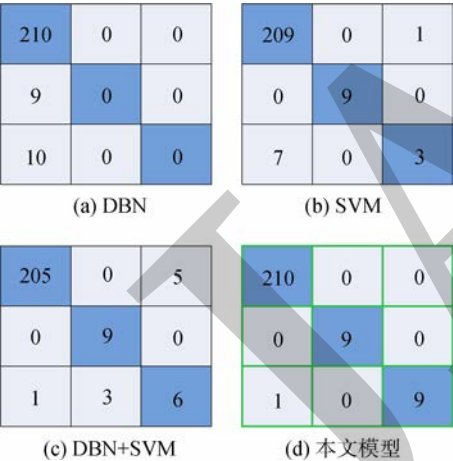


图 6 混淆矩阵
Fig. 6 Confusion matrix

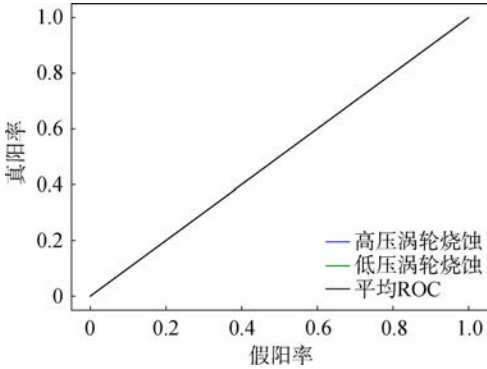


图 7 DBN 模型平均 ROC
Fig. 7 Averaged ROC of DBN model

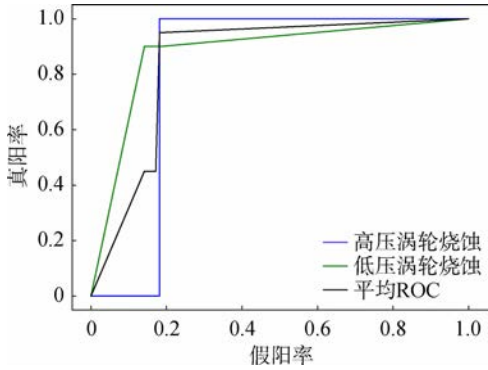


图 8 SVM 模型平均 ROC
Fig. 8 Averaged ROC of SVM model

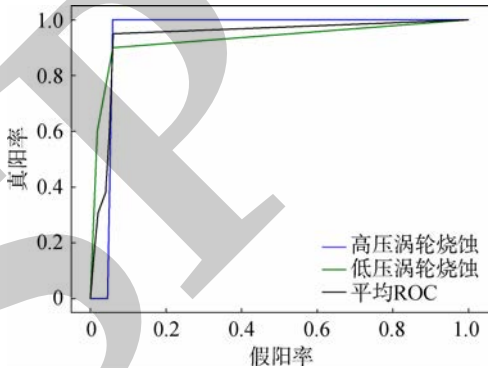


图 9 DBN+SVM 模型平均 ROC
Fig. 9 Averaged ROC of DBN+SVM model

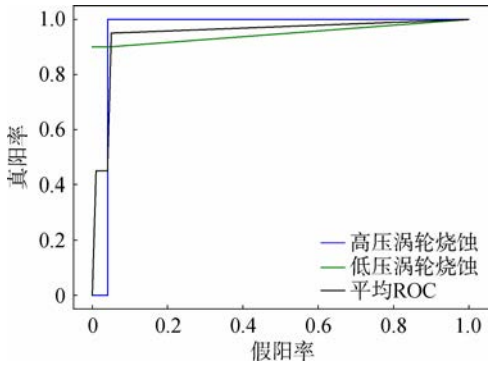


图 10 本文模型平均 ROC
Fig. 10 Averaged ROC of proposed model

4 结 论

针对民航发动机数据样本不平衡问题,本文提出了一种基于不平衡样本驱动的民航发动机气路故障诊断模型。实验通过某航空公司 CFM56-7B 系列的发动机数据样本对模型进行了验证,利用重构误差与训练时间对 DBN 网络结构进行选取,利用准确率与 ROC(AUC 值)评估本文模型

对不均衡样本分类的能力。通过高压涡轮烧蚀与低压涡轮烧蚀两类故障进行诊断发现,本文提出的模型诊断准确率高达 0.996, AUC 值高达 0.948。通过以上实验,总结如下:

1) 本文提出的基于不均衡样本驱动的民航发动机气路故障诊断模型可以有效处理高维特征、样本不均衡问题;

2) 本文模型充分利用 DBN 的特征提取能力,进行民航发动机样本的深度特征挖掘,提取内部特征;

3) 本文模型采用下采样与集成技术,充分利用已有样本中的信息,可以实现多类别故障诊断,有较好的鲁棒性;

4) 方法简单方便,有效地解决了对故障数据的诊断失误,具有一定的实际应用价值。

参考文献:

[1] 李一波,张宗杰,蒋丽英. 航空发动机故障诊断信息融合方法研究现状[J]. 航空发动机,2009,35(3):36-39.
LI Yibo,ZHANG Zongjie,JIANG Liying. Research status of aeroengine fault diagnosis information fusion method [J]. Aeroengine,2009,35(3):36-39. (in Chinese)

[2] 郝英,孙健国,白杰. 航空燃气涡轮发动机气路故障诊断现状与展望[J]. 航空动力学报,2003,18(6):753-760.
HAO Ying,SUN Jianguo,BAI Jie. State of the art and prospect of aircraft engine fault diagnosis using gas path parameters[J]. Journal of Aerospace Power,2003,18(6):753-760. (in Chinese)

[3] 张经璞. 航空发动机状态监控与典型故障分析[D]. 沈阳:沈阳航空航天大学,2017.
ZHANG Jingpu. Aeroengine condition monitoring and typical failure analysis[D]. Shenyang:Shenyang Aerospace University,2017. (in Chinese)

[4] 崔文斌. 航空发动机气路性能估计方法对比研究[D]. 南京:南京航空航天大学,2014.
CUI Wenbin. Aeroengine gas path performance estimation method comparative study[D]. Nanjing:Nanjing University of Aeronautics and Astronautics,2014. (in Chinese)

[5] ZHANG Jing,YANG Yingping,ZHANG Jinmin. A MEC-BP-AdaBoost neural network-based color correction algorithm for color image acquisition equipments[J]. Optics,2016,127(2):776-780.

[6] MONJEZI M,BAHRAMIA. Prediction of blast-induced ground vibration using artificial neural networks[J]. Tunneling and Underground Space Technology,2011,26(1):46-50.

[7] 李爱. 航空发动机磨损故障智能诊断若干关键技术研究[D]. 南京:南京航空航天大学,2013.
LI Ai. Some key technologies for intelligent diagnosis of aero engine wear failure[D]. Nanjing:Nanjing University

of Aeronautics and Astronautics,2013. (in Chinese)

[8] 崔智全. 民航发动机气路参数偏差值挖掘方法及其应用研究[D]. 哈尔滨:哈尔滨工业大学,2013.
CUI Zhiquan. Research on the method of mining parameters deviation of civil aviation engine and its application [D]. Harbin: Harbin Institute of Technology,2013. (in Chinese)

[9] SIMANI S. Residual generator fuzzy identification for automotive diesel engine fault diagnosis[J]. International Journal of Applied Mathematics and Computer Science,2013,23(2):419-438.

[10] 赵光权,葛强强,刘小勇,等. 基于 DBN 的故障特征提取及诊断方法研究[J]. 仪器仪表学报,2016,37(9):1946-1953.
ZHAO Guangquan,GE Qiangqiang,LIU Xiaoyong,et al. Fault feature extraction and diagnosis based on DBN[J]. Chinese Journal of Scientific Instrument,2016,37(9):1946-1953. (in Chinese)

[11] GHOSH A,MAJUMDAR S K. Modeling blast furnace productivity using support vector machines[J]. International Journal of Advanced Manufacturing Technology,2011,52(9/10/11/12):989-1003.

[12] 崔建国,吴灿,董世良,等. 主元分析法和模糊积分的航空发动机气路状态监测[J]. 火力与指挥控制,2014,39(10):130-134.
CUI Jianguo,WU Can,DONG Shiliang,et al. Analysis of aero-engine gas path conditions using principal component analysis and fuzzy integral[J]. Fire Control and Command Control,2014,39(10):130-134. (in Chinese)

[13] 夏丽莎,杨玉英,方华京. 基于 EasyEnsemble 的化工过程故障诊断性能改进[J]. 控制理论与应用,2017,34(2):49-53.
XIA Lisha,YANG Yuying,FANG Huajing. Performance improvement of fault diagnosis in chemical process based on EasyEnsemble[J]. Journal of Control Theory and Applications,2017,34(2):49-53. (in Chinese)

[14] 何玲. 面向故障诊断的异构特征融合与在线不均衡分类研究[D]. 河南 新乡:河南师范大学,2017.
XIANG Ling. Research on heterogeneous features fusion and online imbalance classification for bearing fault diagnosis[D]. Xinxiang Henan:Henan Normal University,2017. (in Chinese)

[15] 陶新民,郝思媛,张冬雪,等. 不均衡数据分类算法的综述[J]. 重庆邮电大学学报(自然科学版),2013,25(1):101-110,121.
TAO Xinmin,HAO Siyuan,ZHANG Dongxue,et al. Overview of classification algorithms for unbalanced data [J]. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition),2013,25(1):101-110,121. (in Chinese)

[16] LECUN Y,BENGIO Y,HINTON G. Deep learning[J]. Nature,2015,521(14539):436-444.

[17] HINTON G E. Training products of experts by minimizing contrastive divergence[J]. Neural Computation,2002,14(8):

1771-1800.

[18] VERBIEST N,RAMENTOL E,CORNELIS C,et al. Pre-processing noisy imbalanced datasets using SMOTE enhanced with fuzzy rough prototype selection[J]. Applied Soft Computing Journal,2014,22:511-517.

[19] 杨宏晖,王芸,孙进才,等. 融合样本选择与特征选择的 AdaBoost 支持向量机集成算法[J]. 西安交通大学学报, 2014,48(12):63-68.

YANG Honghui,WANG Yun,SUN Jincai,et al. AdaBoost support vector machines integration algorithm for fusion sample selection and feature selection[J]. Journal of Xi'an Jiaotong University,2014,48(12):63-68. (in Chinese)

[20] 从威. 非均衡分类的集成学习应用研究[D]. 南京:南京信息工程大学,2017.

CONG Wei. Non-equilibrium classification of integrated learning applications[D]. Nanjing: Nanjing University of Information Science and Technology,2017. (in Chinese)

[21] 殷锴,钟诗胜,那媛,等. 基于 BP 神经网络的航空发动机故障检测技术研究[J]. 航空发动机,2017,43(1):53-57.

YIN Kai,ZHONG Shisheng,NA Yuan,et al. Analysis of aeroengine fault detection based on BP neural network[J]. Aeroengine,2017,43(1):53-57. (in Chinese)

(编辑:陈 越)