

doi: 10.7690/bgzd.2020.12.014

面向发动机故障诊断精度的深度随机森林优化研究

贺鹏程

(海军装备部, 四川 绵阳 621900)

摘要: 为解决装备故障检测存在故障数据少, 难以应用深度学习方法提高性能的问题, 提出一种基于优化深度随机森林的提高装备故障诊断精度方法。根据数据集数据的特点, 构建重采样的样本集合决策树模型, 通过对故障数据中的连续数值进行 C4.5 决策树离散化处理, 使用扩张卷积方法扩展滑动窗口的感测范围, 对训练模型进行实验验证。实验结果表明, 运用深度随机森林的方法对提高装备故障诊断有一定借鉴作用。

关键词: 装备故障诊断; 深度学习; 深度随机森林; 扩张卷积

中图分类号: TP311 **文献标志码:** A

Research on Optimization of Deep Random Forest for Engine Fault Diagnosis Accuracy

He Pengcheng

(Naval Equipment Department, Mianyang 621900, China)

Abstract: In order to solve the problem of insufficient data in equipment fault detection and difficulty to apply deep learning methods for improving performance, a method for improving equipment fault diagnosis accuracy based on optimized deep random forest is proposed. According to the characteristics of the dataset acquiring data, a re-sampled sample set decision tree model was constructed. The continuous value in the fault data was discretized by the C4.5 decision tree, and the extended convolution method was used to expand the sensing range of the sliding window, and finally training the model was verified by experiments. The experimental results show that the method of using deep random forest can be used as a reference for improving equipment fault diagnosis.

Keywords: equipment fault diagnosis; deep learning; deep random forest; expansion convolution

0 引言

随着新装备的不断列装, 装备规模越来越大, 复杂程度越来越高, 对于装备维修人员的要求也不断提升, 给装备诊断工作提出了新的要求, 即如何利用已有的装备运行数据快速检测出装备故障, 完成装备排故。传统的故障诊断方法有支持向量机^[1]、模糊系统、人工神经网络^[2]、决策树^[3]等。由于原有信息化水平等客观原因, 这些方法大多数只是基于少量数据, 利用人为划定的故障特征检测数据差异实现异常数据的判别, 从而对其故障进行分类, 不可避免地导致故障诊断系统的泛化性不强, 难以快速完成小概率故障的诊断。近年来, 随着深度学习理论不断发展, 已经在图像分类、语音识别、大数据分析上取得较大突破^[4]。2018年, 基于深度学习思想的随机森林改进即深度随机森林方法被提出。这是一种通过级联机制构建多层的随机森林, 在小样本量的分类任务中取得了较好的成绩, 并已经被成功应用于离网预测、图像分类等任务^[5-6]。

鉴于装备故障诊断实际情况, 笔者以航空发动机故障数据为实现背景, 提出一种基于深度随机森林的装备故障诊断方法, 通过对深度随机森林进行 C4.5 决策树离散化连续变量优化和膨胀窗口机制改进, 从而实现特征更好的自动提取、异常故障的自动归类, 完成航空发动机的故障诊断。

1 深度随机森林

随机森林是一种统计学习理论, 利用 bootstrap 重采样方法从原始样本集合中不断采样多个样本, 并对每个重采样的样本集合进行决策树建模, 通过投票机制或者平均机制对多个决策树作出的判断进行综合, 实际上就是决策树的深层信息提取。深度随机森林是在这种思路的基础上, 通过叠加多个随机森林构成的一种深度学习模型, 其模型结构如图 1 所示。

该模型具有以下 3 个特征:

1) 使用随机森林形成级联森林, 由随机森林处理初始特征, 获得决策结果的概率向量, 然后与初

收稿日期: 2020-07-20; 修回日期: 2020-08-20

作者简介: 贺鹏程(1985—), 男, 四川人, 硕士, 工程师, 从事军事装备技术及管理、军工产品质量控制及管理、军用标准应用及管理研究。
E-mail: marshallplan@163.com。

始特征组合，传递给下一个随机森林；

2) 利用多粒度扫描机制获取初始特征的时间关系，算法使用不同的滑动窗口扫描不同步编号的原始数据，获得不同维度的初始特征，用于随机森林的训练；

3) 将模型的最终分类结果与森林最后一层的概率结果进行平均，然后取最大值以获得最终结果。

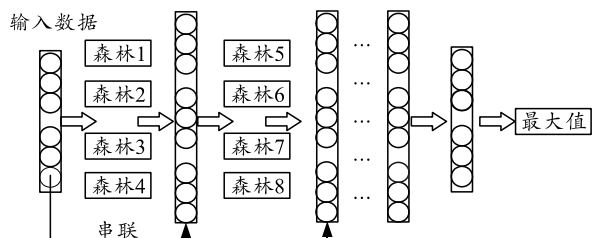


图 1 深度随机森林模型结构

实验结果证明，深度森林在常规数据集上与图像分类、面部识别、情感分类和低维数据上的测试表现类似。与传统的深度学习算法相比，深度森林具有超参数少、模型训练快速收敛、模型训练的计算量小的优势。深度随机森林的主要超参数是每层森林的数量以及每个森林中决策树的数量，要配置的超参数比神经网络小得多，并且，深度随机森林的主要操作是特征信息的获取。该计算比深度神经网络中的大规模矩阵运算和卷积计算要低得多，并且不需要微调参数。

2 装备故障诊断方法

2.1 数据分析

实验选用的数据为美国国家航空航天局 (national aeronautics and space administration, NASA) 提供的涡扇发动机降解模拟数据集 (collection of turbofan engine degradation simulation data, TEDS)。该数据集在操作条件和故障模式的不同组合下，模拟了 4 种不同故障场景，记录多个传感器通道数据来描述故障演化，即从发动机正常状态到最终发动机故障停止运转^[8]。

该数据集由多个时间序列组成。数据包含 26 列，由空格分隔，分别为发动机编号、时间周期、操作设置、传感器测量值等。该数据集具有以下特点：

1) 数据集中的特征类型很多，包括离散值和连续值；

2) 数据的特征量很小，并且不同特征之间存在耦合关系，需要消除和减少这些关系。

该数据集的特点与真实装备数据特点基本相

似，因此，选用该数据集作为模型验证与模型改进的数据对象。

2.2 模型优化

根据数据集的特点，对深度随机森林的优化主要表现在以下 2 点。

1) 通过 C4.5 决策树离散化连续变量。

针对 TED 数据集中的连续特征值，笔者采用 C4.5 决策树对其进行连续值特征的离散化处理。根据连续变量的大小，将需要处理的样本从小到大进行排序。假设该属性对应于总共 N 个属性值，则总共有 $N-1$ 个可能的候选分割值点，并且每个候选分割阈值点的值都在上述排序的属性值中。

2) 基于膨胀窗的特征融合机制的构建。

对于 TED 数据集中不同特征之间的耦合关系，需要对其进行剔除和缩减^[9]。传统的深度随机森林采用连续的滑动窗口来提取输入数据的维数。该方法考虑了特征之间的时序关系，并具有某些特征组合功能，但是缺乏用于不连续特征之间信息组合的处理能力。文献[10]提出的扩张卷积方法用于扩展滑动窗口的感测范围，从而提高模型组合特征的能力。扩大后的窗口操作如图 2 所示，其中空白 1 表示扩大后的大小为 1。

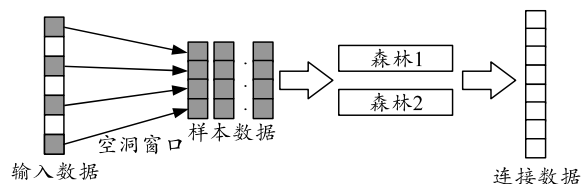


图 2 基于膨胀窗的特征融合机制

根据模型构建与优化过程，该故障诊断系统的训练过程如下：

1) 训练数据，通过多尺度扩张窗口获得不同维度的随机训练样本；

2) 根据获取的训练样本生成 2 个不同的随机森林，一个是完全随机森林，另一个是固定随机森林；

3) 根据随机森林分类的结果，将样本属于某个类别的概率作为训练数据的新维数添加到上层随机森林训练数据中，并连续生成随机森林；

4) 继续执行 2)、3)，直到验证集的准确性降低，即诊断精度达到过拟合与欠拟合的临界点。

3 实验验证

为了测试深度模型性能与优化方法的可行性，在航空发动机故障数据集中选取部分未训练数据用

于实验测试，该测试包括 2 个实验，实验内容与目的如下：

1) 深度模型性能测试实验，对于深层结构和浅层结构性能，使用浅层模型(例如随机森林，决策树和支持向量机)和基于人工神经网络的深度学习模型来预测深度随机森林模型的故障诊断准确率与虚警概率；

2) 深度随机森林超参数测试实验，对深度随机森林中膨胀窗口尺寸与森林层数超参数设置不同数值，进行对比实验，其中膨胀窗口比例与文献[11-12]一致。

将 2 个实验均使用每个类别下的 F1 得分平均值作为度量。计算过程如下：对于每个用户包类别，统计预测正确答案的数量(true positive, TP)、预测错误类别的数量(false positive, FP)，然后计算每个类别下的精度和召回率，最终计算每个类别的 F1 分数以获得均值^[13]。表 1 为实验 1 的实验结果。

表 1 不同模型 F1 值比较 %

模型	F1 数值
随机森林	75.21
C4.5 决策树	70.32
SVM	74.11
深度随机森林	80.90
深度信念网络	81.20

从表中可以看出：使用深度随机森林和深层信念网络获得的 F1 平均值最高，其次是随机林和 C4.5 决策树。这表明：浅层结构模型在处理装备故障诊断等复杂关联问题的能力上不如深度学习模型显著；同时，随机森林通过构建更高级别的决策机制可提高模型适应复杂问题的能力；与浅层模型相比，深度学习模型的深度随机森林和深层信仰网络，具有更好的特征学习能力，F1 值最高。鉴于设置深度信念网络超参数的难度较低，训练耗时较短，使深度随机森林优势明显。

图 3 显示了实验 2 的实验结果。从图 3(a)中可以看出：采用膨胀窗口滑动机制的深度随机森林在很大程度上优于传统的深度随机森林，尤其是在窗口扩大大小为 5 的情况下。考虑到训练数据部分存在强耦合的特征，当最大扩张大小为 5 时，模型达到最高 F1 值，但是随着膨胀窗口的空腔增大，模型的 F1 值减小。其原因是，扩张的尺寸过大会使模型训练的特征尺寸过小，从而导致过度拟合的现象。同样的原因，也导致图 3(b)中随着深度随机森林中的层数增加，F1 分数没有增加反而减少的现象。

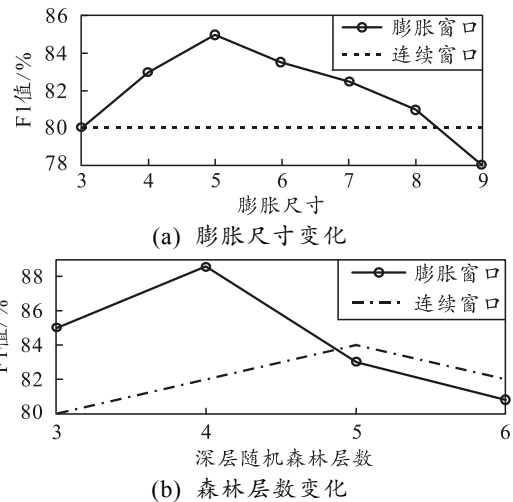


图 3 不同膨胀参数的深度随机森林 F1 值比较

4 结论

笔者使用深度随机森林作为模型进行装备故障的诊断，鉴于 TED 数据集特点，采用连续特征离散化和扩张式滑动窗口机制对森林进行优化。实验结果表明，通过构建深度随机森林可以有效提高故障诊断的精度和降低虚警概率。同时，笔者引入的深度随机森林优化方法也较好地改进深度随机森林的性能。下一步，将继续针对不同问题特征优化深度随机森林，并借鉴深度随机森林设计思想，设计其他实用的深度学习模型和装备故障诊断模型。

参考文献：

- [1] 杨乐, 包磊. 装备故障的时空共现模式挖掘[J]. 兵工自动化, 2016, 35(6): 46-51.
- [2] 陈秀芳, 许利亚, 刘晓春, 等. 决策树算法在装备故障检测中的应用[J]. 兵工自动化, 2015, 34(10): 81-84.
- [3] 何元利, 龚晓峰. 基于卷积神经网络的无线电信号搜索[J]. 兵工自动化, 2017, 36(10): 88-92.
- [4] 肖雄, 董鹏. 基于模糊证据认知图的装备维修风险评估方法[J]. 兵工自动化, 2020, 39(2): 55-58.
- [5] 赵星贺, 周斌, 封会娟, 等. 基于蒙特卡罗仿真评价车辆装备预防性维修间隔期方法[J]. 兵工自动化, 2019, 38(11): 92-96.
- [6] ZHOU Z H, FENG J. Deep Forest: Towards An Alternative to Deep Neural Networks[J]. ArXiv, 2017, 1702: 08835.
- [7] MAJUMDER N, PORIA S, GELBUKH A, et al. Deep Learning-Based Document Modeling for Personality Detection from Text[J]. IEEE Intelligent Systems, 2017, 32(2): 74-79.
- [8] 陈彦彤, 李雨阳, 陈伟楠, 等. 基于深度语义分割的遥感图像海面舰船检测研究[J]. 仪器仪表学报, 2019, 20(15): 1-7.

- [9] XIE C, ZENG W, LU X. Fast Single Image Super-Resolution via Deep Network with Component Learning[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2018, 54(23): 89-95.
- [10] SIDIKE P, KRIEGER E, ALOM M Z, et al. A fast single-image super-resolution via directional edge-guided regularized extreme learning regression[J]. Signal, Image and Video Processing, 2017, 11(5): 961-968.
- [11] CHEN H, ZHUO L, TAO H Y. Study on Building Extraction from High Spatial Resolution Images Using

MMBI[J]. Remote Sensing Technology and Application, 2016, 31(5): 930-938.

- [12] 张浩, 窦奇伟, 栾桂凯, 等. 基于空洞卷积的快速背景自动更换[J]. 计算机应用, 2018, 38(2): 405-409.
- [13] 张义忠, 冯振声. 基于神经网络的复杂电子装备故障诊断系统的仿真研究[J]. 计算机仿真, 2017(4): 39-42.
- [14] 胡翔飞, 王崇亮, 吴厚秀, 等. 暂态零序电流比较方法在配网线路故障诊断分析中的应用[J]. 电力系统装备, 2019(12): 122-123.

(上接第 37 页)

根据式(7)计算属性权重向量得到 $\omega=(0.48,0.21,0.31)^T$ 。

3) 根据式(8)–(11)确定正、负理想方案:

$$B^+ = \{ \{ \langle [0.5, 0.7], [0.5, 0.7] \rangle, \langle [0.1, 0.2], [0.1, 0.2] \rangle \}, \\ \{ \langle [0.4, 0.6], [0.4, 0.6] \rangle, \langle [0.1, 0.2], [0.1, 0.2] \rangle \}, \\ \{ \langle [0.5, 0.6], [0.5, 0.6] \rangle, \langle [0.1, 0.2], [0.1, 0.2] \rangle \} \};$$

$$B^- = \{ \{ \langle [0.3, 0.6], [0.3, 0.6] \rangle, \langle [0.2, 0.3], [0.2, 0.3] \rangle \}, \\ \{ \langle [0.2, 0.5], [0.2, 0.5] \rangle, \langle [0.3, 0.4], [0.3, 0.4] \rangle \}, \\ \{ \langle [0.2, 0.4], [0.2, 0.4] \rangle, \langle [0.2, 0.3], [0.2, 0.3] \rangle \} \}.$$

4) 求取各方案与正、负理想方案之间的加权 Hamming 距离:

$$D_1^+ = 0.0324, \quad D_2^+ = 0.0343, \quad D_3^+ = 0.0371;$$

$$D_1^- = 0.0406, \quad D_2^- = 0.0362, \quad D_3^- = 0.0285.$$

5) 计算各方案与正理想方案的相对贴近度 C_i :

$$C_1=0.45, \quad C_2=0.49, \quad C_3=0.57.$$

根据 C_i 大小进行排序有 $b_1 > b_2 > b_3$, 即 1 号制导雷达的威胁等级最高; 3 号搜索警戒雷达的威胁等级最低。

5 结束语

笔者提出了区间对偶犹豫模糊熵的公理化定

义, 构造了区间对偶犹豫模糊熵公式, 并针对属性权重未知的情况, 给出了一种基于区间对偶犹豫模糊熵的目标威胁评估方法。验证结果表明: 该方法增加了权重的科学性, 提高了评估的准确性, 进一步丰富了模糊多属性决策理论。

参考文献:

- [1] 刘乾, 单洁. 反辐射无人机作战目标威胁评估[J]. 火力与指挥控制, 2014, 39(4): 63-67.
- [2] 郭辉徐, 浩军, 刘凌. 基于区间数 TOPSIS 法的空战目标威胁评估[J]. 系统工程与电子技术, 2014, 31(12): 2914-2917.
- [3] JU Y B, LIU X Y, YANG S H. Interval-valued dual hesitant fuzzy aggregation operators and their applications to multiple attribute decision making[J]. Journal of Intelligent & Fuzzy Systems, 2014, 27: 1203-1218.
- [4] 张建厂, 王文庭. 反辐射无人机攻击目标威胁度分析[J]. 空军雷达学院学报, 2007, 21(3): 11-13.
- [5] 闫菲菲. 对偶犹豫模糊集的熵及其在多属性决策中的应用[D]. 曲阜: 曲阜师范大学, 2015.
- [6] 郭效芝. 模糊不确定性度量的探讨及拓展[D]. 西安: 西北大学, 2004.
- [7] 谷阳阳. 基于信息熵的遥测数据质量维度量化方法[J]. 兵工自动化, 2019, 38(12): 42-45.
- [8] 郭今戈, 卫鸿婧. 集群无人机通信控制方式[J]. 兵器装备工程学报, 2019, 40(11): 122-126.