

控制与决策

Control and Decision

基于多标签学习的旋转机械分级复合故障诊断

马鑫, 陈庆, 柴榕敏, 崔明亮, 王友清

引用本文:

马鑫, 陈庆, 柴榕敏, 崔明亮, 王友清. 基于多标签学习的旋转机械分级复合故障诊断[J]. *控制与决策*, 2022, 37(7): 1772–1778.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2021.0067>

您可能感兴趣的其他文章

Articles you may be interested in

标签Petri网的路径信息在故障诊断中的应用

Application of path information of labeled Petri nets in fault diagnosis

控制与决策. 2021, 36(2): 325–334 <https://doi.org/10.13195/j.kzyjc.2019.0698>

基于广义主成分分析的重构故障子空间建模方法

Reconstructed fault subspace modelling method based on generalized principal component analysis

控制与决策. 2021, 36(4): 808–814 <https://doi.org/10.13195/j.kzyjc.2019.0818>

基于核典型相关性-熵成分分析的工业过程质量监测方法

A quality monitoring method for industrial process based on kernel canonical correlation-entropy component analysis

控制与决策. 2021, 36(12): 2999–3006 <https://doi.org/10.13195/j.kzyjc.2020.0548>

高超声速飞行器间歇故障改进自适应容错控制

Improved adaptive fault-tolerant control of intermittent faults in hypersonic flight vehicle

控制与决策. 2021, 36(11): 2627–2636 <https://doi.org/10.13195/j.kzyjc.2020.0483>

基于社交网络的双知识表达分类方法

Double knowledge representations based classification method from perspective of social networks

控制与决策. 2020, 35(11): 2653–2664 <https://doi.org/10.13195/j.kzyjc.2019.0141>

基于多标签学习的旋转机械分级复合故障诊断

马鑫¹, 陈庆¹, 柴榕敏¹, 崔明亮², 王友清^{1,2†}

(1. 北京化工大学 信息科学与技术学院, 北京 100029;

2. 山东科技大学 电气与自动化工程学院, 山东 青岛 266590)

摘要: 传统故障诊断方法大多是针对单一故障类型, 然而在实际工业中多种故障会同时出现, 即复合故障. 针对复合故障诊断问题, 一些学者引入多标签学习思想, 多标签 K 近邻算法 (ML-KNN) 就是其中之一. 然而 ML-KNN 算法作为一阶算法, 只考虑标签与对应样本数据间的关系, 却忽略了标签间的联系. 针对该问题提出一种分级多标签学习算法, 名为分层多标签 K 近邻算法 (HML-KNN). HML-KNN 算法将机械设备的退化阶段和故障类型分为两级, 将第 1 级得到的标签信息进行转化, 转化后的信息作为新特征放入第 2 级进行判断. HML-KNN 算法是一种高阶算法, 考虑了全局的标签信息, 并在算法中包含了标签的特征转化, 使得到的结果准确率更高. 最后通过 XJTU-SY 数据集验证 HML-KNN 算法在处理复合故障诊断问题上的优越性.

关键词: 多标签学习; ML-KNN; 复合故障; 故障诊断; 分级处理; 相似性搜索

中图分类号: TP273

文献标志码: A

DOI: 10.13195/j.kzyjc.2021.0067

引用格式: 马鑫, 陈庆, 柴榕敏, 等. 基于多标签学习的旋转机械分级复合故障诊断 [J]. 控制与决策, 2022, 37(7): 1772-1778.

Hierarchical compound fault diagnosis of rotating machinery based on multi-label learning

MA Xin¹, CHEN Qing¹, CHAI Rong-min¹, CUI Ming-liang², WANG You-qing^{1,2†}

(1. College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China;

2. College of Electrical and Automation, Shandong University of Science and Technology, Qingdao 266590, China)

Abstract: Traditional fault diagnosis methods are mostly for a single fault type at one time, but in the actual industry, many kinds of faults will occur at the same time, that is compound fault. For the problem of compound fault diagnosis, some scholars have introduced the method of multi-label learning, and the multi-label K -nearest neighbor (ML-KNN) algorithm is one of them. However, as a first-order algorithm, the ML-KNN algorithm only considers the relationship between the label and the corresponding sample data, but ignores the relationship between the labels. In this study, a hierarchical multi-label learning algorithm is proposed, named hierarchical multi-label K -nearest neighbor (HML-KNN). The HML-KNN algorithm categorizes the degradation state as the first level and fault type of machinery as the second level. The first level label information is transformed, and the transformed information is put into the second level as new features for judgment. The HML-KNN algorithm is a high-level algorithm that takes into account the global label information, and includes the feature conversion of the label, which makes the result more accurate. Through the verification on the XJTU-SY bearing data set, the superiority of the HML-KNN algorithm in dealing with compound fault diagnosis is demonstrated.

Keywords: multi-label learning; ML-KNN; compound fault; fault diagnosis; hierarchical processing; similarity search

0 引言

旋转机械不断朝着大型化、复杂化、高速化和自动化方向发展, 这对设备的运行安全提出了更高的要求. 轴承是旋转机械中极其重要的零件之一, 它可以引导轴做旋转运动, 也可以起到减少摩擦的作用. 一

旦运行中轴承等零件出现故障, 将造成巨大的经济损失和安全威胁^[1-3], 因此旋转机械的故障诊断技术一直是学者们研究的重点^[4-5]. 文献[6]对旋转机械的数据收集、指标构建、状态划分和寿命预测做了详尽的总结; 文献[7]基于多元统计故障诊断的理论研究和

收稿日期: 2021-01-12; 录用日期: 2021-04-21.

基金项目: 国家自然科学基金项目 (61822308); 山东省自然科学基金项目 (JQ201812); 青岛市创业创新领军人才计划项目 (19-3-2-4-zhc).

†通讯作者. E-mail: wang_youqing@mail.buct.edu.cn.

工程应用的进展进行了综述.传统的旋转机械故障诊断方法多针对于单一故障,即设备在同一时刻只发生一种故障.然而实际工业过程中,机械设备多为复杂系统,故障的出现具有随机性、继发性、并发性、爆发性、隐蔽性等特质.面对诸多不确定性因素,单一故障诊断存在着很大的局限性.因此,复合故障诊断逐渐成为了研究的新热点^[8-10].

复合故障定义为同时存在两个及以上的故障类型.常见方法如支持向量机^[11]、经验模态分解^[12-13]、神经网络^[14-15]等.一些学者在这些方法的基础上进行了改进和结合,进一步提升了诊断效果.如文献[16]将基于小波的支持向量机与粒子群算法结合,进而优化了分类器的参数,达到更高的准确性;文献[17]提出了辛几何模态分解,将时间序列分解为一组独立的模式分量,可以有效地重构现有模态并消除噪声;文献[18]将多个传感器的数据进行加权处理,再用卷积神经网络进行分类,并验证了算法的有效性.在当前研究中,不少复合故障诊断的方法是对单一故障诊断的扩展,比如将复合故障数据与单一故障数据逐一对比,找出构成复合故障的组合形式.然而,故障的叠加未必是线性的,故障与故障间也存在耦合性.随着设备退化程度的加深,新故障也可能会随之出现.因此故障与故障间的关系和退化程度与故障之间的关系,都是复合故障诊断所需的重要信息.

为了解决这一难题,本文将多标签学习思想应用于复合故障诊断中.多标签学习是针对同时拥有多个标签的样本进行学习,文献[19]对多标签学习算法进行了详细的综述;文献[20]将其应用于电池系统的可靠性管理.除此之外对标签学习也常用于图像识别、文本分类和基因分析领域^[21-22].将样本的标签序列对应到复合故障中,简化了复杂的计算,也进一步挖掘了故障间的关系.

本文通过对ML-KNN算法进行改进,得到一种高阶算法,名为分层多标签 K 近邻算法(hierarchical multi-label K -nearest neighbor),简称HML-KNN. HML-KNN将数据分为退化状态和故障类型两级,并分别进行多标签分类,并将第1级得到的退化标签转化为每组样本的分位数作为个性化的新特征,再将新特征加入原特征中扩充样本维度,放入第2级的模型中进行学习.在XJTU-SY滚动轴承加速寿命试验数据集上的测试^[23]表明,HML-KNN算法在准确率等诸多指标上都体现出良好的性能,从而验证了算法的优越性.

本文的主要创新点有3个:

1)对于多标签学习采用分级处理:将从第1级得到的标签信息进行转化后,作为新特征放入第2级的训练模型,即利用退化状态的标签信息提升对于故障类型判断的准确率.与不进行分级处理的ML-KNN算法相比,准确率的提升可以理解为通过进一步挖掘数据信息,找到了退化阶段与故障间的深层关系,进而提升算法性能.

2)采用创新的标签信息转化方法:第1级得到的标签信息为原始标签组合,如“-1,-1,1”.本文将标签组信息转化为该组数据对应的分位数,即“-1”转化为下四分位数,“1”转化为上四分位数.这样处理后,每组数据得到的新特征都是不同的,可最大程度地满足每组数据的个性化需求,也防止过拟合或欠拟合等问题的出现.

3)信息阶次升级:ML-KNN是一阶算法,忽略与其他标签的关联性;改进后的HML-KNN算法是高阶算法,考虑多个标签之间的关联.准确率的提升体现了理解标签间关系的重要价值,以及算法解读数据全局信息的能力.

1 算法介绍

1.1 多标签学习

传统的监督学习多指样本仅含有一个标签,如判断设备故障与否的二分类问题,判断设备存在某种单一故障的多分类问题,这些都属于单标签学习.而真实的工业情况则更加复杂,设备可能同时存在多个故障,随着退化程度的加深,故障的程度和种类也可能随之变化.随着单标签学习无法完全满足需求,多标签学习的思想应运而生.

多标签学习中一个样本拥有多个标签,即一个标签序列.这个标签序列共同定义了样本的属性.常见的多标签学习分类方法有两种:第1种是按照解决问题的思路划分,第2种是按照标签间的关系划分.

多标签学习也可以按照解决问题的思路分为两大类:问题转化和算法适应.问题转化思想强调将多标签问题转化为已有的问题,如转化为多分类问题或多个二分类问题.算法适应思想则是改变传统算法的结构以适应多标签问题.

多标签学习还可以按照标签间的关系分为3种策略:一阶策略忽略自身与其他标签的关系,如二值相关性(binary relevance)算法、ML-KNN算法^[24];二阶策略考虑标签间的两两关系,如校准标签排序(calibrated label ranking)算法;而高阶策略则是考虑多个标签间的关系,如分类器链(classifier chains)算

法. 多标签学习阶次相关性如图1所示. 在图1中高阶策略的算法关注所有标签间的关系, 掌握全局信息, 更适合处理具有耦合性的复合故障问题.

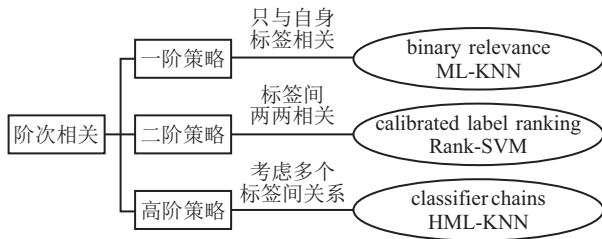


图1 多标签学习阶次相关性

1.2 ML-KNN

ML-KNN算法是一种由KNN算法发展而来的惰性学习算法. 有悖于具有训练建模过程的主动学习, 惰性学习不预先对已知样本进行学习建模, 而是等新样本加入后再分析其与已知样本的关系, 典型的惰性学习算法就是KNN算法. KNN算法通过计算欧氏距离找到 K 个最近邻样本, 再将最近邻中多数类的标签定义为未知样本的标签. ML-KNN则是在KNN的基础上进行改进, 并不是简单的“少数服从多数”, 而是通过贝叶斯准则和最大后验概率计算得到新标签.

定义 X 为一个样本序列, $N(X)$ 为由 X 的 K 个近邻组成的近邻集合. 事件 H_1^l 表示 X 中含有标签 l , 事件 H_0^l 表示 X 中不含有标签 l , 事件 $E_j^l (j \in \{0, 1, \dots, K\})$ 表示 X 的 K 个近邻中含有 j 个标签 l . 那么根据最大后验准则可以得到目标函数

$$y_X(l) = \arg \max_{b \in \{0, 1\}} P(H_b^l) P(E_j^l | H_b^l), \quad (1)$$

进而对样本 X 是否含有标签 l 进行判断. $P(H_b^l)$ 表示 X 是否含有标签 l 的先验概率, $P(E_j^l | H_b^l)$ 表示已知样本标签属性, 其最近邻中含有 j 个标签 l 的概率. $P(H_b^l)$ 和 $P(E_j^l | H_b^l)$ 均可从训练集的样本中计算得到, 最终根据最大后验准则确定测试集中未知样本的标签属性.

相比KNN算法, ML-KNN算法考虑了先验概率, 更擅长处理数据不平衡问题. 同时, ML-KNN算法是一阶算法, 对于每种标签的判断是独立的, 没有充分挖掘标签信息的相关性, 该算法的结构注定会忽略一部分关键信息.

1.3 HML-KNN

HML-KNN的算法结构如图2所示, 将标签序列分为两级, 划分标准可以根据数据的物理意义规定. 例如, 样本特征序列为 $\{6, 47, 49, 15, 42, 41, 7\}$, 含有6个标签, 前3个分为第1级, 后3个分为第2级.

将第1级标签送入贝叶斯分类器, 得到第1级的3个估计标签, 如 $\{1, 1, -1\}$, 将其转化为分位数形式, 即 $\{47, 47, 7\}$, 从而得到新的样本特征序列 $\{6, 47, 49, 15, 42, 41, 7, 47, 47, 7\}$. 将新样本特征序列送入第2级分类器, 得到第2级的估计标签. 将第1级与第2级的估计标签连在一起即得到最终的标签估计序列.

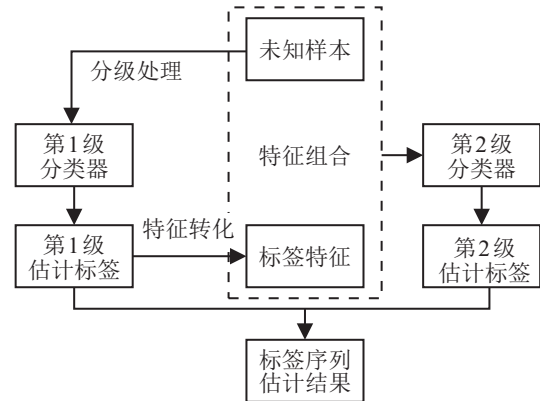


图2 HML-KNN的算法结构

本文提出的HML-KNN算法是对ML-KNN的改进, 主要的改进包括3点:

1) 分级处理: 将原本并行的分类模型进行分级处理, 将第1级分类器得到的估计标签作为新特征放入第2级分类器中, 充分挖掘标签相关性.

2) 特征转化: 将第1级的标签信息转化为标签特征. 放入第2级的标签特征并不是原始标签, 而是该样本序列的分位数, 将标签“-1”转化为下四分位数, 标签“1”转化为上四分位数. 处理后的标签特征是绝对个性化的, 避免了原始标签造成的过拟合或欠拟合等问题.

3) 信息阶次升级: HML-KNN算法是一种考虑了全局标签信息的高阶算法, 第2级的标签与第1级的每个标签都有联系. 多标签学习的内核思想就是挖掘标签间的关联性, 提升结果的可解释性. 高阶算法相比一阶算法, 用到了更多的数据信息, 也给出了更多的启示.

HML-KNN算法作为高阶算法挖掘了标签间更深层次的关联性. 尤其当标签信息对于样本的描述角度不同时, 分级处理将更有助于提升算法的准确率和可解释性.

2 数据描述

本文选择XJTU-SY滚动轴承加速寿命试验数据集作为实验对象^[23]. XJTU-SY数据集采集了轴承从正常到故障状态的振动信号, 包括单一故障和复合故障, 覆盖了不同的故障位置, 记录了完整的全寿命周期数据. 轴承数据一共15个, 依工况分为3组, 每组5

个,如表1所示.

表1 XJTU-SY滚动轴承加速寿命试验数据集

操作条件	轴承数据	轴承寿命	故障位置
工况1 (35 Hz/12 kN)	bearing 1_1	123 min	外圈
	bearing 1_2	161 min	外圈
	bearing 1_3	158 min	外圈
	bearing 1_4	122 min	保持架
	bearing 1_5	52 min	内圈、外圈
工况2 (37.5 Hz/11 kN)	bearing 2_1	491 min	内圈
	bearing 2_2	161 min	外圈
	bearing 2_3	533 min	保持架
	bearing 2_4	42 min	外圈
	bearing 2_5	339 min	外圈
工况3 (40 Hz/10 kN)	bearing 3_1	4 938 min	外圈
	bearing 3_2	2 496 min	内圈、滚珠、保持架、外圈
	bearing 3_3	371 min	内圈
	bearing 3_4	1 515 min	内圈
	bearing 3_5	114 min	外圈

本文将每组的前3个划为训练集,后两个划为测试集,即9个作为训练集,6个作为测试集. 根据经验,将振幅1.5 g作为第1阈值点划分健康阶段和退化阶段,振幅3.5 g作为第2阈值点,划分退化阶段和严重阶段,如图3所示. 故障与退化存在直接关系,当轴承处于健康阶段时,不存在故障;而当轴承处于退化阶段和严重阶段时,则必然存在一种或多种故障. 因此,轴承的退化信息对故障诊断有指导意义.

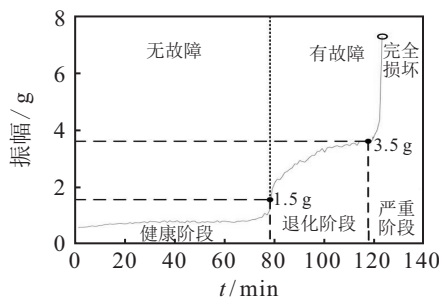


图3 轴承故障与退化的关系

数据预处理分为标签划分和特征提取两项工作. 在标签划分工作中,根据数据的特点将全周期数据分为3个退化阶段,分别为健康阶段、退化阶段、严重阶段,对应前3个标签,分为第1级;将故障位置分为外圈、内圈、保持架,对应后3个标签,分为第2级. 由于滚珠数据样本过少,无法准确验证分类效果,忽略此故障信息. 例如数据处于退化阶段并在外圈和内圈均发生故障,则其对应的标签序列为{-1, 1, -1, 1, 1, -1}. 在特征提取工作中,本文选择了8个常见的时序特征作为样本特征,即

$$p_1 = \frac{\sum_{n=1}^N x(n)}{N}, \quad (2)$$

$$p_2 = \left(\frac{\sum_{n=1}^N \sqrt{|x(n)|}}{N} \right)^2, \quad (3)$$

$$p_3 = \sqrt{\frac{\sum_{n=1}^N (x(n))^2}{N}}, \quad (4)$$

$$p_4 = \max |x(n)|, \quad (5)$$

$$p_5 = \sqrt{\frac{\sum_{n=1}^N (x(n) - p_1)^2}{N - 1}}, \quad (6)$$

$$p_6 = \frac{\sum_{n=1}^N (x(n) - p_1)^3}{(N - 1)p_2^3}, \quad (7)$$

$$p_7 = \frac{\sum_{n=1}^N (x(n) - p_1)^4}{(N - 1)p_2^4}, \quad (8)$$

$$p_8 = \frac{p_4}{p_3}. \quad (9)$$

其中: $p_1 \sim p_4$ 表示时域振动幅值和能量的大小, $p_5 \sim p_8$ 表示时域信号的时间序列分布情况^[25]. 也就是说,送入第1级分类器的原始特征为8个,得到的标签特征为3个,将二者组合后,送入第2级分类器的新特征为11个,加入标签特征后信息变得更加丰富.

为找到最合适的相似样本,本文通过计算相关距离找出15个最近邻建立个性化训练集. 相比欧氏距离,相关距离更能体现数据序列间的相关程度.

3 实验分析

实验部分将HML-KNN算法与ML-KNN算法进行对比,同时也展示了选择相关距离和转化标签信息的优越性. 通过分析3组实验结果验证HML-KNN的优越性.

3.1 评价指标

评价指标分为子分类器评价和整体评价. 其中子分类器评价为每个分类器的准确率(acc_i);整体评价包括4个,分别是:子集准确率(subset_acc)、平均准确率(average_acc)、平均精度(average_prec)、海明损失(hloss). 公式如下所示:

$$acc_i(h) = \frac{|Y_i \cap h(x_i)|}{|Y_i \cup h(x_i)|}, \quad (10)$$

$$subset_acc(h) = \frac{1}{p} \sum_{i=1}^p [[h(x_i) = Y_i]], \quad (11)$$

$$\text{average_acc}(h) = \frac{1}{p} \sum_{i=1}^p \frac{|Y_i \cap h(x_i)|}{|Y_i \cup h(x_i)|}, \quad (12)$$

$$\text{averagez_prec}(h) = \frac{1}{p} \sum_{i=1}^p \frac{|Y_i \cap h(x_i)|}{|h(x_i)|}, \quad (13)$$

$$\text{hloss}(h) = \frac{1}{p} \sum_{i=1}^p \frac{1}{q} |h(x_i) \Delta Y_i|. \quad (14)$$

其中: $h(x_i)$ 是估计标签序列, Y_i 是真实标签序列, q 是所有标签的总个数, Δ 是异或计算; acc_i 是每个标签的分类准确率; subset_acc 是一种更严格的准确率, 要求估计标签序列与真实标签序列完全相同才

算正确; average_prec 体现标签的排序能力, 越大越好; hloss 用来评估一个样本被误分的情况, 越小越好.

子分类器评价和整体评价可以从不同的角度评估算法的性能, 因此本文将使用这两大类指标综合评价算法.

3.2 实验1: 与ML-KNN算法进行对比

实验1将HML-KNN与ML-KNN进行对比, 同样选择相关距离搜索最近邻样本. 区别在于HML-KNN具有分级处理和标签信息转化的步骤, 而ML-KNN则是完全并行地进行分类. 实验结果如表2所示.

表2 HML-KNN与ML-KNN的评价指标对比

	子分类器评价					整体评价				
	acc_1	acc_2	acc_3	acc_4	acc_5	acc_6	subset_acc	average_acc	average_prec	hloss
HML-KNN	0.9968	0.9872	0.9913	0.9309	0.9203	0.9451	0.8993	0.9619	0.9683	0.0381
ML-KNN	0.9968	0.9872	0.9913	0.9277	0.9226	0.9382	0.8974	0.9606	0.9715	0.0394

从表2中可以看出, 第1级的前3个标签准确率是相同的, 区别只存在于后3个标签中. 后3个标签里外圈故障和保持架故障都是HML-KNN算法的准确率高于ML-KNN的准确率, 即对子分类器的评价中, HML-KNN有5个指标更好, 而ML-KNN只有4个. 这也直接导致了HML-KNN的子集准确率和平均准确率较ML-KNN分别高出0.19%和0.13%. 从整体评价指标来看, 虽然HML-KNN的平均精度较ML-KNN低了0.32%, 但HML-KNN有3个指标优于ML-KNN, 有1个指标劣于ML-KNN, 所以认为HML-KNN无论从子分类器评价还是综合评价来分析, 其性能都更加突出.

分级处理对算法的准确率要求更高, 因为第1级的预测结果会直接影响第2级的预测结果. 分级处理是一把“双刃剑”. 若第1级的效果不够突出, 则无论是第2级的准确率还是子集准确率都会降低. 若能够

充分挖掘第1级的退化信息, 则第2级中分类的准确率便会得到相应的提升, 进而提升算法的整体性能.

HML-KNN算法也有助于进一步解读数据信息. 传统故障诊断通常不考虑退化程度, 只关注故障情况. 而实际工业中, 故障的发生必然存在设备的退化, 二者密不可分. 当既考虑退化程度又考虑故障类型时, 可以为实际操作人员对于设备的健康管理提供更多的参考信息.

3.3 实验2: 标签信息转化方式对比

实验2关注标签信息的转化问题, 即对第1级预测得到的3个标签进行特征提取. 特征提取是数据处理的关键环节, 只有将数据转化为分类器更易于理解的信息, 才能提升算法性能. 本实验将标签信息转化为四分位数并使用原始标签作为第2级中的新特征, 通过对比验证HML-KNN算法中标签信息转化这一步骤的必要性. 实验结果如表3所示.

表3 不同标签信息转化方式的评价指标对比

	子分类器评价					整体评价				
	acc_1	acc_2	acc_3	acc_4	acc_5	acc_6	subset_acc	average_acc	average_prec	hloss
四分位数	0.9968	0.9872	0.9913	0.9309	0.9203	0.9451	0.8993	0.9619	0.9683	0.0381
原始标签	0.9968	0.9872	0.9913	0.9167	0.9176	0.9432	0.8864	0.9588	0.9663	0.0412

从表3中可以看出, 第1级的准确率没有区别, 第2级的准确率都是进行了标签信息转化的分类器要高于使用原始数据的分类器. 因此子集准确率和平均准确率较ML-KNN分别高出1.29%和0.31%. 从第2级的准确率更可以看出, 选择四分位数作为转化

后的特征, 大大提升了算法性能. 无论是对子分类器的评价, 还是整体评价, 使用四分位数作为特征优势都更加明显. 因为每个样本的四分位数都是不同的, 这样的转化规则最大程度地满足了样本的个性化需求, 避免了由于原始估计标签造成的过拟合或欠拟合

问题.

工业大数据的运用更容易找到运行的规律,而数据的个性化处理则是对规律的精雕细琢. 算法细节的提升离不开个性化的数据处理过程.

3.4 实验3: 最近邻搜索方法对比

实验3对比了在最近邻搜索过程中使用不同距离算法的结果,即将相关距离与欧氏距离的搜索效果进行对比. 实验结果如表4所示.

表4 不同最近邻搜索方法的评价指标对比

	子分类器评价					整体评价				
	acc_1	acc_2	acc_3	acc_4	acc_5	acc_6	subset_acc	average_acc	average_prec	hloss
相关距离	0.9968	0.9872	0.9913	0.9309	0.9203	0.9451	0.8993	0.9619	0.9683	0.0381
欧氏距离	0.9931	0.9890	0.9922	0.8681	0.9235	0.9286	0.8457	0.9491	0.9493	0.0509

在表4中,选择欧氏距离搜索最近邻的方法,对第1级的3个分类器更有优势,但是整体优势并不明显. 在第2级中分类器4对外圈故障的判断显示出欧氏距离搜索故障样本的明显劣势,较相关距离在此项的准确率低了6.28%. 这也导致了在整体评价里的子集准确率的比较中,选择欧氏距离的方法较相关距离低了5.36%,因此所有整体评价指标都受到影响. 由此表明,选择相关距离搜索最近邻样本较传统的欧氏距离更有优势,更能体现样本序列的相关程度.

HML-KNN 算法是由KNN算法多次改进而来的,也依旧存在着K值选择和距离计算的问题. 这是个性化样本集构建的难点,也是未来算法的改进方向之一.

3.5 实验小结

本文通过3组实验全方位验证了HML-KNN算法的优越性. HML-KNN的分级处理将退化信息加入故障诊断中,充分挖掘了二者间的关系. 通过分类器估计得到的信息,能够提升估计的准确率,这不仅仅挖掘了数据中的信息,也挖掘了算法模型的信息,是模型对数据的理解力的体现. 标签转化则是将第1级得到的退化标签以分位数的形式转化为个性化标签特征,对原始数据特征进行扩维,丰富了数据信息. 最近邻样本的选择是对算法的进一步优化. 作为高阶算法,HML-KNN的全局信息捕捉和解读能力都十分优秀,未来可进一步提升算法细节,如最近邻的个数与搜索方式.

4 结论

为克服传统故障诊断中只考虑单一故障的局限性,本文将多标签学习思路引入复合故障诊断中. 通过改进一阶的ML-KNN算法,提出了一种可以同时判别退化程度和故障类型的高阶算法,名为HML-KNN. HML-KNN通过对标签的分级处理和对特征的个性化转化,不仅兼顾了对退化程度和故障类型的

判断,还通过深度挖掘标签信息提升了算法性能,为实际操作人员提供了更多设备的故障预测和健康管理信息. 通过在XJTU-SY滚动轴承加速寿命试验数据集上的实验,多角度验证了算法的优越性.

未来,可从两个方向继续提升算法性能. 第一,退化程度和故障发生是同时存在的,其中的关联性有待进一步深入挖掘;第二,复合故障诊断满足实际工业的需求,多标签学习思想在工业中的应用可以与机械的物理特征进一步结合,为机械打造数据驱动和设备机理相结合的个性化模型.

参考文献(References)

- [1] Si X S, Wang W B, Hu C H, et al. Remaining useful life estimation — A review on the statistical data driven approaches[J]. European Journal of Operational Research, 2011, 213(1): 1-14.
- [2] 董洁, 游培航, 彭开香. 基于动态内部主元分析和隐马尔科夫模型的动态过程故障检测与分类方法[J]. 控制理论与应用, 2020, 37(10): 2073-2080.
(Dong J, You P H, Peng K X. Dynamic processes fault detection and classification based on dynamic-inner principal component analysis and hidden Markov model[J]. Control Theory & Applications, 2020, 37(10): 2073-2080.)
- [3] 樊继聪, 王友清, 秦泗钊. 联合指标独立成分分析在多变量过程故障诊断中的应用[J]. 自动化学报, 2013, 39(5): 494-501.
(Fan J C, Wang Y Q, Qin S Z. Combined indices for ICA and their applications to multivariate process fault diagnosis[J]. Acta Automatica Sinica, 2013, 39(5): 494-501.)
- [4] Xia M, Li T, Xu L, et al. Fault diagnosis for rotating machinery using multiple sensors and convolutional neural networks[J]. IEEE/ASME Transactions on Mechatronics, 2018, 23(1): 101-110.
- [5] Lei Y G, Li N P, Guo L, et al. Machinery health prognostics: A systematic review from data acquisition to RUL prediction[J]. Mechanical Systems and Signal

- Processing, 2018, 104: 799-834.
- [6] Zhao M H, Zhong S S, Fu X Y, et al. Deep residual shrinkage networks for fault diagnosis[J]. IEEE Transactions on Industrial Informatics, 2020, 16(7): 4681-4690.
- [7] Wang Y Q, Si Y B, Huang B, et al. Survey on the theoretical research and engineering applications of multivariate statistics process monitoring algorithms: 2008-2017[J]. The Canadian Journal of Chemical Engineering, 2018, 96(10): 2073-2085.
- [8] Teng W, Ding X, Cheng H, et al. Compound faults diagnosis and analysis for a wind turbine gearbox via a novel vibration model and empirical wavelet transform[J]. Renewable Energy, 2019, 136: 393-402.
- [9] 张坤鹏, 姜斌, 陈复扬, 等. 基于时变模型辨识的高速列车复合故障诊断[J]. 控制与决策, 2019, 34(2): 274-278.
(Zhang K P, Jiang B, Chen F Y, et al. Time-varying model identified based coupled fault diagnosis for high speed trains[J]. Control and Decision, 2019, 34(2): 274-278.)
- [10] Yi Z G, Pan N, Guo Y. Mechanical compound faults extraction based on improved frequency domain blind deconvolution algorithm[J]. Mechanical Systems and Signal Processing, 2018, 113: 180-188.
- [11] Ji D Z, Pan H Y, Cheng J S. Rolling bearing fault detection and diagnosis based on composite multiscale fuzzy entropy and ensemble support vector machines[J]. Mechanical Systems and Signal Processing, 2017, 85: 746-759.
- [12] Jiang H K, Li C L, Li H X. An improved EEMD with multiwavelet packet for rotating machinery multi-fault diagnosis[J]. Mechanical Systems and Signal Processing, 2013, 36(2): 225-239.
- [13] Lei Y G, Lin J, He Z J, et al. A review on empirical mode decomposition in fault diagnosis of rotating machinery[J]. Mechanical Systems and Signal Processing, 2013, 35(1/2): 108-126.
- [14] Huang R Y, Liao Y X, Zhang S H, et al. Deep decoupling convolutional neural network for intelligent compound fault diagnosis[J]. IEEE Access, 2019, 7: 1848-1858.
- [15] Jia F, Lei Y G, Lu N, et al. Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization[J]. Mechanical Systems and Signal Processing, 2018, 110: 349-367.
- [16] Liu Z W, Cao H R, Chen X F, et al. Multi-fault classification based on wavelet SVM with PSO algorithm to analyze vibration signals from rolling element bearings[J]. Neurocomputing, 2013, 99: 399-410.
- [17] Pan H Y, Yang Y, Li X, et al. Symplectic geometry mode decomposition and its application to rotating machinery compound fault diagnosis[J]. Mechanical Systems and Signal Processing, 2019, 114: 189-211.
- [18] Shan P F, Lv H, Yu L M, et al. A multisensor data fusion method for ball screw fault diagnosis based on convolutional neural network with selected channels[J]. IEEE Sensors Journal, 2020, 20(14): 7896-7905.
- [19] Zhang M L, Zhou Z H. A review on multi-label learning algorithms[J]. IEEE Transactions on Knowledge and Data Engineering, 2014, 26(8): 1819-1837.
- [20] Li S H, Cao H L, Yang Y P. Data-driven simultaneous fault diagnosis for solid oxide fuel cell system using multi-label pattern identification[J]. Journal of Power Sources, 2018, 378: 646-659.
- [21] 刘慧婷, 冷新杨, 王利利, 等. 联合嵌入式多标签分类算法[J]. 自动化学报, 2019, 45(10): 1969-1982.
(Liu H T, Leng X Y, Wang L L, et al. A joint embedded multi-label classification algorithm[J]. Acta Automatica Sinica, 2019, 45(10): 1969-1982.)
- [22] Zhu P F, Xu Q, Hu Q H, et al. Multi-label feature selection with missing labels[J]. Pattern Recognition, 2018, 74: 488-502.
- [23] Wang B, Lei Y G, Li N P, et al. A hybrid prognostics approach for estimating remaining useful life of rolling element bearings[J]. IEEE Transactions on Reliability, 2020, 69(1): 401-412.
- [24] Zhang M L, Zhou Z H. ML-KNN: A lazy learning approach to multi-label learning[J]. Pattern Recognition, 2007, 40(7): 2038-2048.
- [25] Lei Y G. Fault diagnosis based on novel hybrid intelligent model[J]. Chinese Journal of Mechanical Engineering, 2008, 44(7): 112.

作者简介

马鑫(1995-), 女, 博士生, 从事预测与健康管理、智能故障诊断和人工智能算法的研究, E-mail: 2019400171@mail.buct.edu.cn;

陈庆(1998-), 男, 硕士生, 从事复杂工程监测的研究, E-mail: 827769616@qq.com;

柴榕敏(1996-), 女, 硕士生, 从事深度学习的研究, E-mail: 1079722614@qq.com;

崔明亮(1994-), 男, 硕士生, 从事基于深度学习的故障诊断的研究, E-mail: qq2196055414@163.com;

王友清(1981-), 男, 教授, 博士生导师, 从事容错控制、状态监测、生物医学过程的建模与控制以及迭代学习控制等研究, E-mail: wang_youqing@mail.buct.edu.cn.

(责任编辑: 闫妍)