

doi: 10.3969/j.issn.2095-0411.2022.06.007

基于生物免疫机理的连续学习故障诊断方法

李 栋¹, 龚兰兰¹, 刘树林², 孙 欣², 雷 勇³

(1. 常州大学 石油工程学院, 江苏 常州 213164; 2. 上海大学 机电工程与自动化学院, 上海 200444;
3. 渤海船舶重工有限责任公司, 辽宁 葫芦岛 125004)

摘要: 智能故障诊断方法是保障机械设备安全可靠运行的重要手段, 然而, 现有智能故障诊断方法缺乏连续学习能力, 诊断时无法有效识别未参与训练的故障类型。在生物免疫系统智能机理的启发下, 提出一种连续学习故障诊断方法。此方法使用 k 近邻策略优化诊断过程提升运行速度; 通过不断对诊断数据的学习, 实时培养和更新记忆细胞; 利用新类型记忆细胞阈值识别未参与训练的故障类型, 并培养新类型记忆细胞; 利用新记忆细胞阈值更新已知类型记忆细胞, 用于提升故障诊断性能。在一定条件下, 此方法退化为普通的监督学习故障诊断方法。利用 20 个标准数据集和凯斯西储大学轴承故障数据集对此方法进行了评估, 实验结果表明此故障诊断方法具有良好的故障诊断性能。

关键词: 故障诊断; 生物免疫机理; 连续学习; 机器学习

中图分类号: TP 306

文献标志码: A

文章编号: 2095-0411(2022)06-0054-09

Fault Diagnosis Method with Continual Learning Capacity Based on the Biological Immune Mechanism

LI Dong¹, GONG Lanlan¹, LIU Shulin², SUN Xin², LEI Yong³

(1. School of Petroleum Engineering, Changzhou University, Changzhou 213164, China; 2. School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200444, China; 3. Bohai Shipbuilding Heavy Industry Co., Ltd., Huludao 125004, China)

Abstract: Intelligent fault diagnosis methods play a significant role to ensure the safe and reliable operation of mechanical equipment. However, they cannot effectively classify the unseen samples in the testing stage, for they lack continual learning ability. A continual learning fault diagnosis method CLFDMD was proposed, inspired by the intelligent mechanisms of the biological immune system. It used k nearest neighbor rule to optimize the testing process for improving the running speed. It cultivated and updated memory cells by continual learning testing data during the testing stage. The new types

收稿日期: 2022-01-26。

基金项目: 国家自然科学基金资助项目(52075310)。

作者简介: 李栋(1981—), 男, 山西长治人, 博士, 副教授。E-mail: lidong@cczu.edu.cn

引用本文: 李栋, 龚兰兰, 刘树林, 等. 基于生物免疫机理的连续学习故障诊断方法[J]. 常州大学学报(自然科学版), 2022, 34(6): 54-62.

of memory cells were cultivated after they recognized the unseen types of data with the help of their threshold. The memory cells of known types were updated by threshold of new memory cells to improve the fault diagnosis performance. It degenerates into a standard supervised learning fault diagnosis method at certain conditions. To assess the performance and possible advantages of CLFDMD, the experiments on twenty well-known datasets from the UCI repository and the ball bearing test dataset from Case Western Reserve University Bearing Data Center were performed. Results showed that it had better fault diagnosis performance.

Key words: fault diagnosis; biological immune mechanism; continual learning; machine learning

智能故障诊断方法是保障机械设备安全可靠运行的重要手段^[1-4],神经网络^[5-7]、支持向量机^[7-8]、 k 近邻^[9]、深度学习^[10-11]以及迁移学习^[12-13]等方法在机械设备故障诊断中发挥了重要作用,并取得了大量的研究成果。然而,这些方法多属于批量学习方式,即离线学习训练、在线故障诊断。一方面,这些方法在学习训练时假设获取了所有类型的故障数据,事实上很难一次性获取全部类型的故障数据,从而使这些故障诊断方法无法有效识别出未参与训练的故障类型;另一方面,学习训练结束后,学习过程结束,这些故障诊断方法不能在诊断过程中继续学习获取新知识;再者,某些故障数据或故障特征数据会随着机械设备服役时间的延长而发生变化,从而使由某一时间段采集的数据训练得到的故障诊断模型,经过一段时间后其故障诊断性能退化,很难高效开展故障诊断。

定期重新训练是解决上述问题的常用方法,然而重新训练会消耗大量的时间和空间,同时也降低了故障诊断方法的时效性。增量学习故障诊断方法^[14-15]和连续学习故障诊断方法^[16-18]是解决上述问题的有效手段。这类方法以渐进方式进行知识更新,并能对以前的知识修正和加强,使得更新后的知识能适应新的数据,而不必重新对全部数据进行学习的算法。

基于生物免疫机理的等大小记忆细胞连续学习分类方法(Continual learning classification method with constant-sized memory cells based on artificial immune system, C-CLCM),是受生物免疫系统随入侵抗原进化而进化以及接种疫苗增强免疫能力等智能机理启发而提出的^[16]。学习训练阶段培养初级记忆细胞,测试过程中通过不断对测试数据的学习,实时培养记忆细胞。当学到新知识时,通过培养新类型记忆细胞,使C-CLCM能有效识别出未参与训练的样本;当对已学到知识进行更新时,通过培养新的已知类型记忆细胞,提升C-CLCM的分类性能。C-CLCM对已知类型的数据进行分类,对未知类型的数据进行聚类,在一定条件下,C-CLCM退化为普通的监督学习分类方法,实验结果验证了C-CLCM具有良好的分类性能。

C-CLCM通过被抗原所攻击潜在记忆细胞的相邻记忆细胞属性或所攻击记忆细胞及其相邻记忆细胞属性确定抗原类型,因此,寻找(潜在)记忆细胞的相邻记忆细胞是C-CLCM的重要工作之一。然而,C-CLCM寻找(潜在)记忆细胞的相邻记忆细胞的过程繁琐,导致测试阶段C-CLCM运行效率低,难以满足在线学习与测试的要求。因此需要在C-CLCM的基础上进一步对其寻找(潜在)记忆细胞的相邻记忆细胞的过程进行优化,并将其应用于机械设备故障诊断。因此,提出连续学习故障诊断方法(Continual learning fault diagnosis method, CLFDMD)。

1 等大小记忆细胞连续学习分类方法(C-CLCM)

C-CLCM将 n 维 m 类抗体(训练数据)培养的记忆细胞定义为

$$\langle c_1, c_2, \dots, c_n, p, q_1, q_2, \dots, q_m \mid c \in \mathbf{R}, p, q \in \mathbf{N} \rangle \quad (1)$$

式中: c_i ($i=1, 2, \dots, n$) 为记忆细胞第 i 维中心坐标; p 为细胞代数; q_j ($j=1, 2, \dots, m$) 为

记忆细胞中第 j 类细胞核数量。

学习训练时根据抗体在状态空间的分布情况直接生成等大小超立方体记忆细胞，而不是通过细胞分裂的方式生成记忆细胞^[16-18]。因此 C-CLCM 的细胞集中只有记忆细胞，而没有空细胞，从而降低了细胞集的存储空间。

测试阶段，当抗原（测试数据）攻击的是潜在记忆细胞时，由此潜在记忆细胞的相邻记忆细胞属性确定抗原类型；当抗原攻击的是记忆细胞时，由此记忆细胞及其相邻记忆细胞属性确定抗原类型。抗原具体类型由式 (2) 确定，式 (2) 中的参数由式 (3)~式 (6) 确定。

$$a_k = \max (a_1, a_2, \dots, a_m) \quad (2)$$

$$a_j = \frac{2}{\pi} \arctan (k_j) \quad (3)$$

$$k_j = \frac{\frac{q_{1j}}{q_{1\max}d_1} + \frac{q_{2j}}{q_{2\max}d_2} + \dots + \frac{q_{tj}}{q_{t\max}d_t}}{t} \quad (j=1, 2, \dots, m) \quad (4)$$

$$q_{l\max} = \max (q_{l1}, q_{l2}, \dots, q_{lm}) \quad (l=1, 2, \dots, t) \quad (5)$$

$$d_l = \text{Euclidean_distance} (c_a, c_l) \quad (l=1, 2, \dots, t) \quad (6)$$

式中： t 为被抗原所攻击潜在记忆细胞相邻记忆细胞数量，或是被抗原所攻击记忆细胞及其相邻记忆细胞数量； c_a 为抗原中心坐标； c_l ($l=1, 2, \dots, t$) 为第 l 个相邻记忆细胞坐标。

寻找被抗原所攻击（潜在）记忆细胞的相邻记忆细胞是 C-CLCM 在测试阶段的重要工作之一。C-CLCM 首先扩大被抗原所攻击（潜在）记忆细胞，然后寻找是否有记忆细胞覆盖到此细胞，若有，则这些记忆细胞是此细胞的相邻记忆细胞，若没有，则继续扩大被抗原所攻击细胞，直到有记忆细胞覆盖到此细胞。如图 1 所示，记忆细胞 1 扩大 2 次后，被记忆细胞 2 和记忆细胞 3 覆盖，则记忆细胞 2 和记忆细胞 3 是记忆细胞 1 的相邻记忆细胞；记忆细胞 4 扩大 4 次后，被记忆细胞 1、记忆细胞 5 和记忆细胞 6 覆盖，则记忆细胞 1、记忆细胞 5 和记忆细胞 6 是记忆细胞 4 的相邻记忆细胞。由此可以看出，C-CLCM 寻找被抗原所攻击细胞的相邻记忆细胞过程复杂，计算效率低，尤其在多维空间表现更明显。

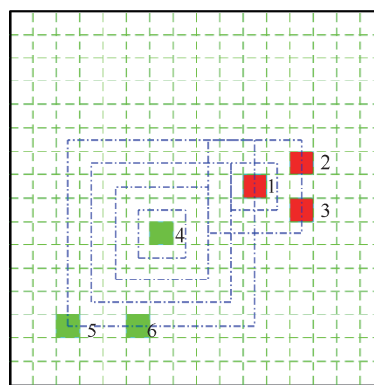


图 1 C-CLCM 寻找相邻记忆细胞过程

Fig.1 The searching adjacent memory cells process of C-CLCM

2 连续学习故障诊断方法 (CLFDMD)

2.1 基本思想

C-CLCM 中抗原的类型，由其所攻击潜在记忆细胞的相邻记忆细胞属性或所攻击记忆细胞及其相邻记忆细胞属性决定。从图 1 可以看出，虽然每个（潜在）记忆细胞的相邻记忆细胞数量不同，但都围绕着被抗原所攻击的细胞，而且每个记忆细胞的大小相同。为此，使用 k 近邻方法寻找被抗原所攻击（潜在）记忆细胞的相邻记忆细胞，可以提高 C-CLCM 寻找相邻记忆细胞的速度。从图 1 可以看出，若 $k=2$ ，记忆细胞 1 的相邻记忆细胞是记忆细胞 2 和记忆细胞 3；记忆细胞 4 的相邻记忆细胞是记忆细胞 1 和记忆细胞 6。

与 C-CLCM 寻找（潜在）记忆细胞的相邻记忆细胞的过程相比，使用 k 近邻方法找到的被抗原

所攻击(潜在)记忆细胞的相邻记忆细胞数量相同,而且会失去(潜在)记忆细胞共边或共点的记忆细胞,尤其是在高维空间。但是在高维空间,记忆细胞数量有限,与(潜在)记忆细胞共边或共点的记忆细胞相对较少,因此,对抗原识别的准确性影响也相对较小。

测试阶段,当抗原攻击的是潜在记忆细胞时,由此潜在记忆细胞的 k 个相邻记忆细胞属性确定抗原类型;当抗原攻击的是记忆细胞时,由此记忆细胞及其 $k-1$ 个相邻记忆细胞属性确定抗原类型,抗原具体类型仍由式(2)确定。当符合条件的记忆细胞数量 $k_i > k$ 时,则从抗原与记忆细胞最大距离相等的记忆细胞中随机舍弃 $k_i - k$ 个记忆细胞。

2.2 训练与诊断过程

CLFDMD 的训练过程与 C-CLCM 相同,其测试过程与 C-CLCM 相似,不同之处仅为寻找相邻记忆细胞的方法不同。为此,CLFDMD 在 C-CLCM 涉及需要赋值 3 个参数(细胞代数 p ,新记忆细胞阈值 a_u ,新类型记忆细胞阈值 a_l)的基础上增加了参数 k 。

CLFDMD 训练时,只涉及参数 p 。首先根据参数 p 生成潜在记忆细胞在每维的中心点坐标,并计算细胞边长,然后依次取抗体培养记忆细胞。假设抗体类型为 j ,若抗体位于某潜在记忆细胞内,则将此潜在记忆细胞进化为记忆细胞,并令其 $q_j = 1$;若抗体位于某记忆细胞内,则令此记忆细胞的 $q_j = q_j + 1$;若抗体位于细胞壁上,则生成多个记忆细胞。

CLFDMD 测试时,涉及参数 a_u , a_l 和 k 。首先找到被抗原所攻击的潜在记忆细胞或记忆细胞,然后找到潜在记忆细胞的 k 个相邻记忆细胞,或记忆细胞的 $k-1$ 个相邻记忆细胞,接着使用式(2)计算 a_k 。当 $a_k > a_u$,若被抗原所攻击的是潜在记忆细胞,则进化为记忆细胞,令其 $q_k = 1$,若被抗原所攻击的是记忆细胞,则更新其属性,令其 $q_k = q_k + 1$,并输出抗原类型 k ;当 $a_k < a_l$,则识别出新类型数据,将记忆细胞结构调整为 $\langle c_1, c_2, \dots, c_n, p, q_1, q_2, \dots, q_m, q_{m+1} \rangle$,并令 $q_{m+1} = 0$;将此潜在记忆细胞进化为记忆细胞,并令其 $q_{m+1} = 1$ 。

为控制新类型数量,或分析新类型数据与已知类型数据之间的关系,与 C-CLCM 一样,最后根据类间相似程度对识别结果进行聚类。

3 仿真与实例分析

为了验证 CLFDMD 的性能与存在的问题,用 UCI 标准数据集和凯斯西储大学的轴承故障数据集,对 CLFDMD 的故障诊断性能进行分析,并与文献中的常见故障诊断方法进行对比。

3.1 UCI 标准数据集仿真分析

3.1.1 分类性能

与 C-CLCM 一样,对已知类型的数据进行分类,对未知类型的数据进行聚类,当 a_u 值很大,且 a_l 值很小时,CLFDMD 退化为普通的监督学习分类方法,因此 CLFDMD 必须具有良好的分类性能。为了验证 CLFDMD 在分类方面的性能与存在问题,用 20 个未丢失样本的 UCI 标准数据集进行验证,见表 1。并与常用的分类方法贝叶斯网络(Naïve Bayesian networks, NB)、决策树 C4.5 算法(C4.5)、 k 近邻算法(k -nearest neighbors, k NN)、RIPPER、BP 神经网络、支持向量机的序列最小最优化方法(Sequential minimal optimization algorithm, SMO)以及 C-CLCM 进行比较。

另外,CLFDMD 与 C-CLCM 一样,可以用抗体整体培养记忆细胞,也可以将抗体分为 n 段,分别培养记忆细胞;将 CLFDMD 整体培养记忆细胞标记为 CLFDMD₁,将 CLFDMD 分段培养记忆细胞标记为 CLFDMD _{n} 。

采用 10 倍交叉验证实验, 将数据集均匀分为互不重叠的 10 份, 轮流将其中 1 份作为测试数据集, 9 份作为训练数据集, 训练模型并完成测试。每个数据集运行 10 次, 结果取平均值。为了对比 CLFDMD 与 C-CLCM 使用不同相邻记忆细胞寻找方法对识别结果的影响, $C-CLCM_1$, $C-CLCM_n$, $CLFDMD_1$ 和 $CLFDMD_n$ 使用完全相同的训练数据和测试数据进行对比, 因此, $C-CLCM_1$ 和 $C-CLCM_n$ 的结果与文献 [16] 略有差异。各算法的计算结果见表 2, 其他算法的计算结果来自文献 [16] 和文献 [19]。从表 2 可以看出, 20 个数据集中, 有 19 个数据集的 $CLFDMD_1$ 的分类准确率高于 $C-CLCM_1$; 此 19 个数据集中, 有 15 个数据集的 $CLFDMD_n$ 的分类准确率高于 $C-CLCM_n$ 。由此可看出, CLFDMD 的分类性能优于 C-CLCM。与 C-CLCM 以及其他经典的分类算法相比, CLFDMD 具有良好的分类性能。

表 1 各数据集简介

| Table 1 Descriptions of datasets | | | | |
|----------------------------------|--------|--------|------|------|
| 序号 | 数据集名称 | 样本数量 | 样本维度 | 类别数量 |
| 1 | 天平 | 625 | 4 | 3 |
| 2 | 糖尿病 | 768 | 8 | 2 |
| 3 | 哈伯曼生存 | 306 | 3 | 2 |
| 4 | 心脏病 | 270 | 13 | 2 |
| 5 | 电离层 | 351 | 34 | 2 |
| 6 | 鸢尾花卉 | 150 | 4 | 3 |
| 7 | 国际象棋 | 3 196 | 35 | 2 |
| 8 | 字母 | 20 000 | 16 | 26 |
| 9 | 淋巴造影 | 148 | 18 | 4 |
| 10 | 和尚问题 1 | 124 | 6 | 2 |
| 11 | 和尚问题 2 | 169 | 6 | 2 |
| 12 | 和尚问题 3 | 122 | 6 | 2 |
| 13 | 托儿所 | 12 960 | 8 | 5 |
| 14 | 图像分割 | 2 310 | 19 | 7 |
| 15 | 声呐 | 208 | 60 | 2 |
| 16 | 垃圾邮件 | 4 601 | 57 | 2 |
| 17 | 车辆 | 846 | 18 | 4 |
| 18 | 元音 | 990 | 13 | 11 |
| 19 | 波形 | 5 000 | 40 | 3 |
| 20 | 葡萄酒 | 178 | 13 | 3 |

表 2 各算法的分类准确率

Table 2 The accuracy of each algorithm in each dataset

%

| 数据集 | 准确率 | | | | | | | | | |
|--------|-------|-------|-------------------|--------|-------|-------|-----------------|-----------------|----------------------|----------------------|
| | NB | C4.5 | 3NN ¹⁾ | RIPPER | BP | SMO | $C-CLCM_1$ | $C-CLCM_n$ | $CLFDMD_1$ | $CLFDMD_n$ |
| 天平 | 90.53 | 77.82 | 86.74 | 80.91 | 85.67 | 87.62 | 89.42($p=2$) | 89.09($p=2$) | 89.33($p=4, k=8$) | 89.55($p=2, k=6$) |
| 糖尿病 | 75.75 | 74.49 | 73.86 | 75.22 | 77.04 | 77.07 | 73.91($p=2$) | 75.25($p=5$) | 74.79($p=5, k=6$) | 76.81($p=5, k=7$) |
| 哈伯曼生存 | 75.06 | 71.05 | 69.77 | 72.72 | 74.20 | 73.40 | 72.06($p=2$) | | 72.91($p=2, k=10$) | |
| 心脏病 | 83.59 | 78.15 | 79.11 | 78.70 | 83.30 | 83.81 | 78.52($p=2$) | 83.63($p=2$) | 81.48($p=2, k=8$) | 84.96($p=3, k=9$) |
| 电离层 | 82.17 | 89.74 | 86.02 | 89.30 | 87.07 | 87.93 | 88.26($p=5$) | 93.05($p=3$) | 87.32($p=6, k=2$) | 93.16($p=4, k=3$) |
| 鸢尾花卉 | 95.53 | 94.73 | 95.20 | 93.93 | 84.80 | 84.87 | 96.73($p=8$) | 96.13($p=4$) | 97.20($p=4, k=6$) | 96.40($p=3, k=2$) |
| 国际象棋 | 87.79 | 99.44 | 96.56 | 99.13 | 98.92 | 95.78 | 84.37($p=3$) | 94.09($p=4$) | 94.65($p=5, k=8$) | 93.76($p=3, k=7$) |
| 字母 | 64.12 | 87.95 | 94.54 | 87.10 | 81.93 | 81.57 | 94.23($p=8$) | 91.26($p=9$) | 96.26($p=10, k=4$) | 91.82($p=8, k=7$) |
| 淋巴造影 | 83.13 | 75.84 | 81.74 | 77.36 | 82.26 | 85.94 | 76.22($p=4$) | 80.74($p=5$) | 83.04($p=2, k=10$) | 81.69($p=8, k=7$) |
| 和尚问题 1 | 73.38 | 80.61 | 78.97 | 83.87 | 97.69 | 79.58 | 75.48($p=6$) | 87.50($p=9$) | 74.76($p=6, k=10$) | 90.56($p=9, k=10$) |
| 和尚问题 2 | 56.83 | 57.75 | 54.74 | 56.21 | 100 | 58.70 | 83.61($p=2$) | 66.15($p=10$) | 84.73($p=2, k=3$) | 66.09($p=7, k=3$) |
| 和尚问题 3 | 93.45 | 92.95 | 86.72 | 84.80 | 88.32 | 93.45 | 78.52($p=3$) | 87.05($p=5$) | 83.44($p=7, k=2$) | 89.75($p=9, k=10$) |
| 托儿所 | 90.30 | 97.18 | 97.36 | 98.67 | 99.43 | 93.02 | 84.09($p=2$) | 91.77($p=6$) | 97.82($p=7, k=7$) | 92.42($p=9, k=8$) |
| 图像分割 | 80.16 | 96.79 | 96.12 | 95.53 | 91.35 | 89.72 | 96.10($p=6$) | 96.48($p=6$) | 97.26($p=8, k=2$) | 96.93($p=8, k=7$) |
| 声呐 | 67.71 | 73.61 | 83.76 | 75.45 | 78.67 | 77.88 | 77.12($p=10$) | 74.71($p=7$) | 82.98($p=4, k=2$) | 79.09($p=4, k=10$) |
| 垃圾邮件 | 79.56 | 92.68 | 89.80 | 92.67 | 91.22 | 90.48 | 89.36($p=7$) | 91.63($p=10$) | 92.12($p=9, k=6$) | 92.56($p=9, k=3$) |
| 车辆 | 44.68 | 72.28 | 70.21 | 68.32 | 81.11 | 74.08 | 68.23($p=4$) | 72.33($p=5$) | 72.35($p=5, k=4$) | 73.10($p=6, k=10$) |
| 元音 | 62.90 | 80.20 | 96.99 | 71.17 | 92.73 | 70.61 | 98.64($p=9$) | 95.82($p=4$) | 99.31($p=6, k=1$) | 95.60($p=5, k=10$) |
| 波形 | 80.01 | 75.25 | 77.67 | 79.14 | 86.56 | 86.94 | 79.75($p=2$) | 82.77($p=3$) | 81.54($p=6, k=10$) | 83.77($p=2, k=10$) |
| 葡萄酒 | 97.46 | 93.20 | 95.85 | 93.14 | 98.02 | 98.76 | 95.51($p=2$) | 97.02($p=5$) | 96.29($p=4, k=7$) | 97.87($p=3, k=5$) |

说明: 1) 3NN 为 k 近邻算法, k 的取值为 3。

3.1.2 参数 k 对 CLFDMD 分类性能的影响

CLFDMD 共涉及 4 个参数, 由于其算法结构与 C-CLCM 相同, 在文献 [16] 中 C-CLCM 详细讨论参数 a_u , a_l 和 p 对 C-CLCM 分类性能的影响, 在此仅讨论参数 k 对 CLFDMD 分类性能的影响。采用 10 倍交叉实验, 训练数据和测试数据相同时, 仅改变参数取值。CLFDMD 的参数 $p=5, 10$, 参数 k 取 1~10。每个数据集运行 10 次, 取平均值, 如图 2 所示。由于 CLFDMD_l 和 CLFDMD_n 算法结构相同, 在此仅分析 CLFDMD_l 的运行结果。从图 2 可看出, 相同条件下, 参数 p 对 CLFDMD 识别准确率的影响很小, 参数 k 的取值对 CLFDMD 识别准确率有一定的影响, 与具体数据集中各类数据在状态空间的具体分布有关, 没有明显的规律性, 目前需要根据经验选择参数 k 。

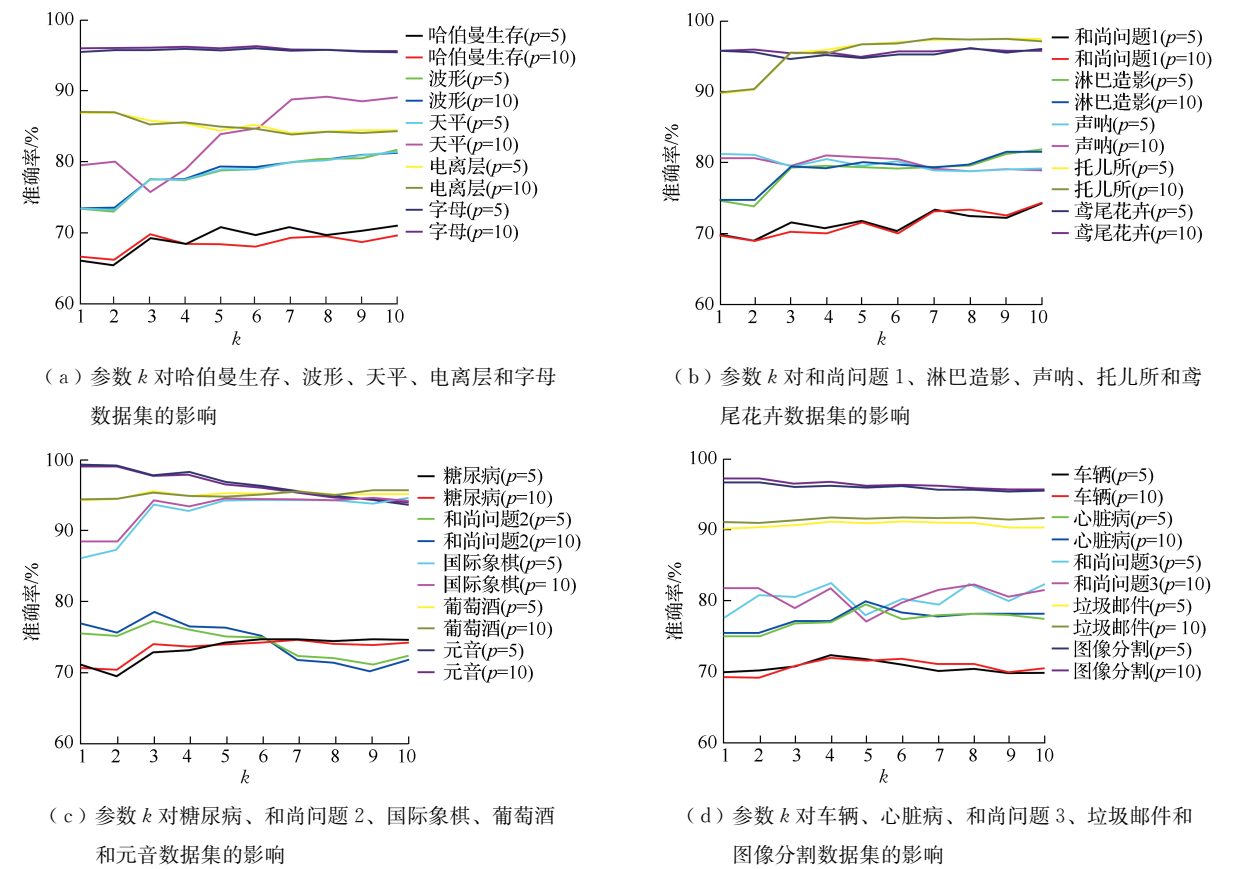


图 2 参数 k 对 CLFDMD 分类性能的影响

Fig.2 Effect of k on the classification ability of CLFDMD

3.1.3 运算速度

由于 C-CLCM 与 CLFDMD 记忆细胞培养过程完全相同, 在此仅对比测试速度。随机选取 20 个标准数据集中 90% 的数据培养记忆细胞, 剩余的数据用于测试。C-CLCM 与 CLFDMD 使用完全相同的训练数据与测试数据, C-CLCM 参数 $p=10$, CLFDMD 参数 $p=10, k=10$ 。Matlab 程序在联想 P710 工作站 (CPU: E5-2643 v4; 内存: 32 G; 操作系统: Windows 10 专业版) 运行。每个数据集运行 10 次, 结果取平均值, 运行结果见表 3。

从表 3 中可看出, 对同一数据集, C-CLCM_l 耗时最长是 CLFDMD_l 的 25 000 多倍 (淋巴造影数据集), 最少是 CLFDMD_l 的 40 多倍 (字母数据集), 平均为 CLFDMD_l 的 3 100 多倍; C-CLCM_n 耗时最长是 CLFDMD_n 的 7 500 多倍 (淋巴造影数据集), 最少是 CLFDMD_n 的 2 倍多 (垃圾邮件数据集), 平均是 CLFDMD_n 的 700 多倍。由此可看出, 与 C-CLCM 相比, CLFDMD 具有很高的运行速度。

表 3 C-CLCM 与 CLFDMD 的运行速度
Table 3 The run time of C-CLCM and CLFDMD

s

| 数据集 | 运行速度 | | | |
|--------|---------------------|---------------------|---------------------|---------------------|
| | C-CLCM ₁ | C-CLCM _n | CLFDMD ₁ | CLFDMD _n |
| 天平 | 37.327 5 | 93.383 3 | 0.064 1 | 0.081 0 |
| 糖尿病 | 15.939 3 | 10.751 9 | 0.056 5 | 0.397 8 |
| 哈伯曼生存 | 0.802 4 | | 0.010 4 | |
| 心脏病 | 168.333 1 | 10.507 4 | 0.018 3 | 0.115 3 |
| 电离层 | 81.337 1 | 100.146 6 | 0.025 4 | 0.403 5 |
| 鸢尾花卉 | 0.523 4 | 1.160 9 | 0.005 4 | 0.051 0 |
| 国际象棋 | 1 950.837 6 | 118.411 7 | 0.908 5 | 24.563 8 |
| 字母 | 1 350.281 8 | 1 537.242 2 | 32.592 5 | 11.367 2 |
| 淋巴造影 | 180.935 5 | 471.947 6 | 0.007 2 | 0.062 7 |
| 和尚问题 1 | 21.253 6 | 69.544 2 | 0.017 5 | 0.033 1 |
| 和尚问题 2 | 27.830 8 | 102.617 9 | 0.013 4 | 0.056 7 |
| 和尚问题 3 | 6.238 1 | 1.896 6 | 0.006 6 | 0.074 6 |
| 托儿所 | 6 567.538 4 | 1 825.838 8 | 14.164 7 | 1.252 7 |
| 图像分割 | 43.507 9 | 18.731 4 | 0.383 8 | 4.344 8 |
| 声呐 | 52.959 4 | 11.674 1 | 0.011 0 | 0.209 8 |
| 垃圾邮件 | 10 244.804 1 | 164.793 9 | 2.242 9 | 56.537 6 |
| 车辆 | 34.666 5 | 13.009 8 | 0.068 6 | 0.846 4 |
| 元音 | 13.027 9 | 15.119 6 | 0.101 8 | 0.959 1 |
| 波形 | 16 807.534 5 | 369.257 3 | 2.450 1 | 68.627 2 |
| 葡萄酒 | 11.996 0 | 8.491 7 | 0.011 8 | 0.130 0 |

此外，除字母数据集和托儿所数据集外，CLFDMD_n 耗时最长约是 CLFDMD₁ 的 28 倍（波形数据集），最少约是 CLFDMD₁ 的 1.26 倍（天平数据集），平均约为 CLFDMD₁ 的 12 倍。字母数据集和托儿所数据集 CLFDMD_n 耗时少于 CLFDMD₁ 的原因是 CLFDMD₁ 培养的记忆细胞数量多于 CLFDMD_n。通常，CLFDMD 培养的记忆细胞数量少于数据集中抗体数量，然而，当抗体位于细胞壁上时，一个抗体可以培养多个记忆细胞，从而使部分数据集培养的记忆细胞数量多于抗体数量。

为了验证这一现象，将每个数据集的所有数据用于培养记忆细胞，参数 $p=10$ ，培养的记忆细胞数量见表 4。从表 4 中可看出，一般情况下 CLFDMD₁ 培养的记忆细胞数量要少于 CLFDMD_n，这是因为 CLFDMD_n 通过分段后，使得抗体的数量是 CLFDMD₁ 的 n 倍，因此培养的记忆细胞数量要多。当有较多的抗体位于细胞壁上时，CLFDMD_n 培养的记忆细胞数量会少于 CLFDMD₁，这是因为抗体只是某些维

表 4 CLFDMD 培养记忆细胞数量
Table 4 The number of memory cells of CLFDMD

| 数据集 | 记忆细胞数量 | |
|--------|---------------------|---------------------|
| | CLFDMD ₁ | CLFDMD _n |
| 天平 | 625 | 500 |
| 糖尿病 | 768 | 5 791 |
| 哈伯曼生存 | 288 | |
| 心脏病 | 557 | 2 424 |
| 电离层 | 350 | 9 919 |
| 鸢尾花卉 | 194 | 706 |
| 国际象棋 | 3 196 | 114 171 |
| 字母 | 18 668 | 14 710 |
| 淋巴造影 | 148 | 344 |
| 和尚问题 1 | 124 | 125 |
| 和尚问题 2 | 169 | 107 |
| 和尚问题 3 | 122 | 732 |
| 托儿所 | 13 254 | 1 358 |
| 图像分割 | 2 307 | 29 889 |
| 声呐 | 217 | 7 723 |
| 垃圾邮件 | 4 130 | 195 374 |
| 车辆 | 846 | 13 380 |
| 元音 | 990 | 11 762 |
| 波形 | 5 000 | 199 993 |
| 葡萄酒 | 178 | 2 314 |

度位于细胞壁, CLFDMD_n 将抗体重组后, 只在少数子集上有抗体位于细胞壁上, 从而使 CLFDMD_n 培养的记忆细胞数量较少。当 CLFDMD₁ 和 CLFDMD_n 培养的记忆细胞数量相当时, 由于在测试阶段, CLFDMD_n 要重复测试过程 n 次, 从而导致 CLFDMD_n 的运行速度较慢。只有当 CLFDMD_n 培养的记忆细胞数量远少于 CLFDMD₁ 时, CLFDMD_n 的运行速度才能快于 CLFDMD₁。

3.2 轴承故障数据集仿真分析

为了进一步验证 CLFDMD 在连续学习故障诊断中的性能, 采用凯斯西储大学的轴承故障数据进行验证, 并与文献 [20] 中的方法进行对比。从 3.1 中的分析可以看出, 通常情况下, CLFDMD₁ 的性能略优于 CLFDMD_n, 因此用 CLFDMD₁ 进行连续学习故障诊断。

将凯斯西储大学的轴承故障数据中风扇端加速度采样数据用 db16 小波包 8 层分解, 选用高频小波系数能量做为特征向量, 生成正常样本 500 个, 生成内圈 0.177 8 mm 故障样本、内圈 0.355 6 mm 故障样本、内圈 0.533 4 mm 故障样本、滚动体 0.177 8 mm 故障样本、滚动体 0.355 6 mm 故障样本、滚动体 0.533 4 mm 故障样本、外圈 0.177 8 mm 故障样本、外圈 0.355 6 mm 故障样本和外圈 0.533 4 mm 故障样本各 50 个, 共计 950 个。

3.2.1 训练样本包括所有类型

从各类样本中随机选取 1 个样本作为训练样本 (共计 10 个训练样本), 其余作为测试样本 (共计 940 个测试样本), 参数 $p=8, k=1, a_u=0.88, a_l=0$ 。运行 100 次, 诊断结果取平均值, 与其他方法的对比结果见表 5。从表 5 中可以看出, CLFDMD 具有最高的诊断准确率。

表 5 训练样本包括所有类型时各种算法的准确率
Table 5 Accuracy of each algorithm when the training samples include all types %

| 算法 | ARIS | CLONALG | BayesNet | J48 | BP | SVM | AHr-detector | CLFDMD |
|-------|-------|---------|----------|-------|-------|-------|--------------|--------|
| 诊断准确率 | 64.41 | 91.54 | 72.84 | 66.28 | 68.05 | 94.67 | 91.37 | 97.63 |

3.2.2 训练样本缺少部分类型

从正常样本和滚动体 0.533 4 mm 故障各随机选取 1 个样本作为训练样本 (共计 2 个训练样本), 其余样本作为测试样本 (共计 948 个测试样本), 参数 $p=8, k=1, a_u=0.88, a_l=0.85$ 。运行 100 次, 诊断结果取平均值, 与其他方法的对比见表 6。从表 6 中可以看出, CLFDMD 具有最高的诊断准确率。这是因为 CLFDMD 具有连续学习能力, 能在诊断过程中识别出未参与训练的故障类型。

表 6 训练样本缺少部分类型时各种算法的准确率
Table 6 Accuracy of each algorithm when the training samples include partial types %

| 算法 | ARIS | CLONALG | BayesNet | J48 | BP | SVM | AHr-detector | CLFDMD |
|-------|-------|---------|----------|-------|-------|-------|--------------|--------|
| 诊断准确率 | 52.63 | 57.78 | 57.62 | 52.61 | 42.82 | 45.17 | 78.90 | 88.28 |

4 结 论

- ① CLFDMD 是受生物免疫系统利用记忆细胞快速识别二次入侵抗原, 以及记忆细胞随入侵抗原进化而进化、接种疫苗增强免疫等智能机理的启发下提出的一种具有连续学习能力的故障诊断方法。② CLFDMD 在诊断过程中不仅能学习新知识, 识别出未参与训练的故障类型, 还能通过对诊断数据的不断学习, 更新已有知识, 提升故障诊断能力。当 a_u 取值较大, 且 a_l 取值较小时, CLFDMD 退化为一般的监督学习故障诊断方法。③ CLFDMD 在 C-CLCM 的基础上, 采用 k 近邻方法寻找被抗原

攻击(潜在)记忆细胞的相邻记忆细胞,与 C-CLCM 相比,显著提高了测试速度。④ CLFDMD 涉及到的参数 p , k , a_u 和 a_l 主要依靠经验取值,仍需进一步研究如何自主优化参数。

参考文献:

- [1] LEI Y G, YANG B, JIANG X W, et al. Applications of machine learning to machine fault diagnosis: a review and roadmap[J]. Mechanical Systems and Signal Processing, 2020, 138: 106587.
- [2] LIU R N, YANG B Y, ZIO E, et al. Artificial intelligence for fault diagnosis of rotating machinery: a review[J]. Mechanical Systems and Signal Processing, 2018, 108: 33-47.
- [3] 雷亚国,贾峰,孔德同,等. 大数据下机械智能故障诊断的机遇与挑战[J]. 机械工程学报, 2018, 54(5): 94-104.
- [4] 王国彪,何正嘉,陈雪峰,等. 机械故障诊断基础研究“何去何从”[J]. 机械工程学报, 2013, 49(1): 63-72.
- [5] LI X, YANG Y, HU N Q, et al. Discriminative manifold random vector functional link neural network for rolling bearing fault diagnosis[J]. Knowledge-Based Systems, 2021, 211: 106507.
- [6] 李巍华,翁胜龙,张绍辉. 一种萤火虫神经网络及在轴承故障诊断中的应用[J]. 机械工程学报, 2015, 51(7): 99-106.
- [7] WANG B, KE H W, MA X D, et al. Fault diagnosis method for engine control system based on probabilistic neural network and support vector machine[J]. Applied Sciences, 2019, 9(19): 4122.
- [8] LI X, YANG Y, PAN H Y, et al. A novel deep stacking least squares support vector machine for rolling bearing fault diagnosis[J]. Computers in Industry, 2019, 110: 36-47.
- [9] LEI Y G, ZUO M J. Gear crack level identification based on weighted K nearest neighbor classification algorithm[J]. Mechanical Systems and Signal Processing, 2009, 23(5): 1535-1547.
- [10] XING S B, LEI Y G, WANG S H, et al. Distribution-invariant deep belief network for intelligent fault diagnosis of machines under new working conditions[J]. IEEE Transactions on Industrial Electronics, 2021, 68(3): 2617-2625.
- [11] 王毅,戴国洪,王克胜. 深度学习技术在预测维修中的应用综述[J]. 常州大学学报(自然科学版), 2019, 31(3): 1-22.
- [12] YANG B, LEI Y G, JIA F, et al. A polynomial kernel induced distance metric to improve deep transfer learning for fault diagnosis of machines[J]. IEEE Transactions on Industrial Electronics, 2020, 67(11): 9747-9757.
- [13] YANG B, LEI Y G, JIA F, et al. An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings[J]. Mechanical Systems and Signal Processing, 2019, 122: 692-706.
- [14] CHAI Z, ZHAO C H. Multiclass oblique random forests with dual-incremental learning capacity[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 31(12): 5192-5203.
- [15] YU W K, ZHAO C H. Broad convolutional neural network based industrial process fault diagnosis with incremental learning capability[J]. IEEE Transactions on Industrial Electronics, 2020, 67(6): 5081-5091.
- [16] LI D, LIU S L, GAO F R, et al. Continual learning classification method with constant-sized memory cells based on the artificial immune system[J]. Knowledge-Based Systems, 2021, 213: 106673.
- [17] LI D, LIU S L, GAO F R, et al. Continual learning classification method with new labeled data based on the artificial immune system[J]. Applied Soft Computing, 2020, 94: 106423.
- [18] 李栋,刘树林,孙欣. 针对不连续时变样本空间的连续学习故障诊断方法[J]. 机械工程学报, 2021, 57(1): 157-167.
- [19] KOTSIANTIS S B, ZAHARAKIS I D, PINTELAS P E. Machine learning: a review of classification and combining techniques[J]. Artificial Intelligence Review, 2006, 26(3): 159-190.
- [20] DONG L, LIU S L, ZHANG H L. A method of anomaly detection and fault diagnosis with online adaptive learning under small training samples[J]. Pattern Recognition, 2017, 64: 374-385.

(责任编辑:谭晓荷)