

## 基于改进 ENN2 聚类算法的多故障诊断方法

文天柱<sup>1</sup>, 许爱强<sup>1</sup>, 程 恭<sup>2</sup>

(1. 海军航空工程学院 科研部, 山东 烟台 264001; 2. 海军装备部, 北京 100841)

**摘要:** 针对可拓神经网络无法解决多故障诊断的问题, 建立问题模型, 将多故障诊断问题转化为多特征样本的聚类问题. 从模型结构和学习算法两个方面对 ENN2 进行改进, 提出基于改进 ENN2 聚类算法的多故障诊断方法, 并对其参数和时间复杂度进行分析. 采用工程实例对所提出的方法进行验证, 结果表明, 所提出的方法能够解决离线多故障诊断问题, 且得到的诊断模型可用于在线状态监控, 具有较好的应用前景.

**关键词:** 多故障诊断; 可拓神经网络; 改进 ENN2 聚类算法; 状态监控

中图分类号: TP18

文献标志码: A

## Multi-fault diagnosis method based on improved ENN2 clustering algorithm

WEN Tian-zhu<sup>1</sup>, XU Ai-qiang<sup>1</sup>, CHNEG Gong<sup>2</sup>

(1. Department of Scientific Research, Naval Aeronautical and Astronautical University, Yantai 264001, China;

2. Equipment Department of Navy, Beijing 100841, China. Correspondent: WEN Tian-zhu, E-mail: wentianzhu1987@aliyun.com)

**Abstract:** For the problem that multi-fault diagnosis can not be solved by the extension neural network, a problem model is built, and the multi-fault diagnosis problem is transformed into the clustering problem for multi-attribute samples. ENN2 is improved from two faces of the model structure and learning algorithm, and the multi-fault diagnosis method based on the improved ENN2 clustering algorithm is proposed with the analysis of parameters and time complexity. The proposed method is verified by an engineering instance. The results show that the method can resolve the offline multi-fault diagnosis problem, and the obtained diagnosis model can also be applied to online fault monitoring, so it has a wide application prospect.

**Keywords:** multi-fault diagnosis; extension neural network; improved ENN2 clustering algorithm; condition monitoring

### 0 引言

故障诊断主要包含故障检测、故障隔离和故障辨识3方面内容, 按照被诊断系统的故障并发个数可以将其分为单故障诊断和多故障诊断. 单故障诊断针对在任意时刻最多只有一个故障发生的单故障系统, 多故障诊断针对在同一时刻可能有多个故障发生的多故障系统<sup>[1-2]</sup>. 由于多故障系统在现实生活中普遍存在, 研究多故障诊断十分必要. 现有的多故障诊断方法主要是将多故障系统转换为单故障系统再进行诊断, 但多故障系统故障机理复杂, 故障之间可能相关也可能无关, 所以转换过程困难, 诊断结果差强人意. 另外, 由于故障预测和健康管理技术的发展, 学者们已经不满足于对确定性故障的诊断, 对潜在故障的早期诊断也成为研究的热点<sup>[3]</sup>.

可拓神经网络是继模糊神经网络、遗传神经网络、进化神经网络等之后又一新型神经网络, 具有神经网络非线性处理能力强和可拓学能够定性、定量地解决矛盾问题的双重优点, 并在分类聚类、模式识别和故障诊断等领域得到应用<sup>[4]</sup>. 可拓神经网络主要包括 ENN1 (extension neural network type1)<sup>[5-6]</sup>和 ENN2<sup>[7]</sup>两种形式, 它们都采用两层网络结构, 输入层和输出层之间通过双权连接, 主要区别在于学习算法不同, ENN1 采用有导师监督的学习算法, ENN2 采用无导师监督的学习算法, 所以 ENN1 常用于解决分类问题, ENN2 常用于解决聚类问题. 可拓神经网络在故障诊断相关领域的应用较多, 文献[3]提出将可拓神经网络用于功率变压器的故障诊断; 文献[8]提出将可拓神经网络用于汽油发动机的失火故障诊断;

收稿日期: 2014-04-23; 修回日期: 2014-08-09.

基金项目: 武器装备预研基金项目(9140A27020212JB14311); “泰山学者”建设工程专项经费项目.

作者简介: 文天柱(1987-), 男, 博士生, 从事复杂装备故障诊断的研究; 许爱强(1963-), 男, 教授, 博士生导师, 从事复杂装备自动测试技术等研究.

文献[9]提出将可拓神经网络用于光电系统的孤岛检测;文献[10]提出将可拓神经网络用于大功率设备的局部放电模式识别;文献[11]提出将遗传算法与可拓神经网络进行融合用于轴承的故障诊断等.上述研究都针对单一目标的故障诊断或模式识别,并不适用于多故障诊断和状态监控,所以需要设计一种多故障诊断方法,使其在能诊断确定性多故障的同时,能够发现早期的潜在故障.

鉴于ENN2在诊断过程中不需要已知故障类型和训练样本,适合工程应用,而且采用可拓距离表示样本到各故障聚类中心的距离,有助于发现潜在故障,本文主要研究基于改进ENN2聚类算法的多故障诊断方法.首先建立问题模型,将多故障诊断问题转化为聚类问题;然后针对在此类聚类问题中样本可以同时属于多个聚类的特殊性,对ENN2进行改进,提出基于改进ENN2聚类算法的多故障诊断方法,并对其参数和时间复杂度进行分析;最后采用工程实例进行验证,结果表明,所提出方法能够解决离线的多故障诊断问题,且得到的诊断模型可用于在线状态监控,具有较好的应用前景.

## 1 问题模型

设待诊断样本集

$$X = \{x_i, i = 1, 2, \dots, m\},$$

样本  $x_i$  可用物元表示为

$$M_i = \begin{bmatrix} x_i & c_1 & v_{i1} \\ & c_2 & v_{i2} \\ & \vdots & \vdots \\ & c_n & v_{in} \end{bmatrix}.$$

其中:  $c_j$  为样本  $x_i$  的第  $j$  个特征;  $v_{ij}$  为  $c_j$  的特征值,  $j = 1, 2, \dots, n$ .

任意待诊断样本物元  $M_i$  ( $i = 1, 2, \dots, m$ ) 所属的故障类型集为  $F_i$ , 如果  $F_i = \emptyset$ , 则表示待诊断样本物元  $M_i$  正常; 如果  $|F_i| = 1$ , 则表示待诊断样本物元  $M_i$  为单故障; 如果  $|F_i| > 1$ , 则表示待诊断样本物元  $M_i$  为多故障, 其中  $|F_i|$  表示故障类型集  $F_i$  中元素的个数.

多故障诊断问题的实质是根据待诊断物元  $M_i$  的特征, 确定故障类型集  $F_i$ , 因此, 可以将其看作是对待诊断样本集  $X$  的聚类过程. 但与传统聚类不同, 多故障诊断问题中每个待诊断物元  $M_i$  可以同时属于多个聚类.

ENN2是一种新型聚类算法, 模型结构如图1所示<sup>[12]</sup>, 包括输入层和输出层, 输入层与样本的特征相对应, 输出层与样本的故障类型相对应, 输入、输出层之间通过双权连接. 由于输出层各神经元之间存在反

馈抑制, 每次聚类只能有一个输出神经元被激活, 无法解决多故障诊断问题, 为此需要对ENN2进行改进.

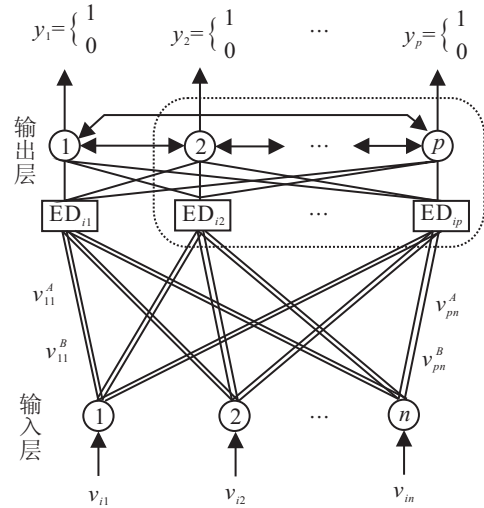


图1 ENN2模型结构

## 2 改进ENN2聚类算法

针对多故障诊断聚类问题的特点, 同时为了解决ENN2距离参数选择困难、聚类效率低等问题, 从模型结构和学习算法两个方面对ENN2进行改进.

### 2.1 模型结构

对模型结构的改进主要包括采用初等关联函数定义可拓距离; 改变连接输入神经元和输出神经元的双权含义; 取消输出层的反馈抑制, 并将输出神经元由二值输出扩展为三值输出.

#### 1) 可拓距离.

ENN2通过计算样本  $M_i$  与各聚类中心  $z_t$  ( $t = 1, 2, \dots, p$ ) 的可拓距离  $ED_{it}$  进行聚类, 可拓距离的表达式为

$$ED_{it} = \sum_{j=1}^n \left\{ \frac{|v_{ij} - u_{tj}| - \lambda}{\lambda} + 1 \right\}.$$

其中:  $v_{ij}$  为样本  $M_i$  的第  $j$  个特征值,  $u_{tj}$  为聚类中心  $z_t$  的第  $j$  个特征值,  $\lambda$  为距离参数. 聚类方法为: 对于任意  $t \in \{1, 2, \dots, p\}$ , 当  $ED_{it} = \min_{t=1, 2, \dots, p} (ED_{it})$  时,  $M_i \in T_t$ , 即样本  $M_i$  属于到聚类中心  $z_t$  可拓距离最小的聚类  $T_t$ .

由于ENN2中定义的可拓距离没有考虑样本特征值的节域, 而且距离参数  $\lambda$  需要根据经验给定, 主观性强, 聚类方法也不适用于多故障诊断, 给出基于关联函数的可拓距离计算公式<sup>[13]</sup>

$$ED_{it} = \min_{j=1, 2, \dots, n} (k(v_{ij})), \quad (1)$$

其中  $k(v_{ij})$  为特征值  $v_{ij}$  关于节域  $V_j = \langle v_j^C, v_j^D \rangle$  和经典域  $V_{0tj} = \langle v_{tj}^A, v_{tj}^B \rangle$  的关联函数.

若节域  $V_j$  和经典域  $V_{0tj}$  有公共端点, 且  $v_{ij}$  取到该公共端点, 则  $k(v_{ij}) = 0$ ; 否则, 若  $\rho(v_{ij}, V_{0tj}) =$

$\rho(v_{ij}, V_j)$  且  $v_{ij} \notin V_{0tj}$ , 则

$$k(v_{ij}) = \frac{\rho(v_{ij}, V_{0tj})}{D(v_{ij}, V_{0tj}, V)} - 1.$$

其他情况下, 有

$$k(v_{ij}) = \frac{\rho(v_{ij}, V_{0tj})}{D(v_{ij}, V_{0tj}, V)}.$$

聚类方法为: 对于任意  $t \in \{1, 2, \dots, p\}$ , 当  $ED_{it} \geq 0$  时,  $M_i \in T_t$ , 样本  $M_i$  属于聚类  $T_t$ ; 当  $0 > ED_{it} \geq \delta > -1$  时,  $M_i \in T_t$ , 样本  $M_i$  可能属于聚类  $T_t$ ; 当  $\delta > ED_{it}$  时,  $M_i \notin T_t$ , 样本  $M_i$  不属于聚类  $T_t$ .

2) 双权.

ENN2采用每个特征经典域  $V_{0tj}$  的上限  $v_{tj}^B$  和下限  $v_{tj}^A$  作为连接输入神经元和输出神经元的双权. 因为每个特征的聚类中心是其经典域的中点, 即  $v_{tj}^A = u_{tj} - \lambda$ ,  $v_{tj}^B = u_{tj} + \lambda$ , 所以当  $\lambda$  取常量时, 采用经典域上、下限为双权的ENN2等价于采用聚类中心为单权的ENN2. 另外, 在ENN2中, 距离参数  $\lambda$  是一个用户自定义的参数, 需要根据工程经验进行选取, 距离参数  $\lambda$  取值的不同直接导致最终聚类结果的不同. 为了解决ENN2距离参数的取值问题, 提出采用聚类中心  $u_{tj}$  和可变距离参数  $\lambda_{tj}$  作为改进ENN2聚类算法的双权, 且有  $v_{tj}^A = u_{tj} - \lambda_{tj}$ ,  $v_{tj}^B = u_{tj} + \lambda_{tj}$ , 通过学习算法可以实现双权的自动更新, 无需人为设定.

3) 输出层.

ENN2的输出层各神经元之间存在反馈抑制, 即对样本集  $X$  聚类时, 只能有一个神经元处于激活状态, 无法实现一个样本同时属于多个聚类. 为此, 取消输出层各神经元之间的反馈抑制, 输出神经元仅根据样本  $M_i$  与聚类中心  $z_t$  的可拓距离  $ED_{it}$  判断输出. 对于任意  $t \in \{1, 2, \dots, p\}$ , 如果  $M_i \in T_t$ , 则输出  $y_t = 1$ ; 如果  $M_i \in T_t$ , 则输出  $y_t = 0$ ; 如果  $M_i \notin T_t$ , 则输出  $y_t = -1$ . 在原来二值输出的基础上扩展为三

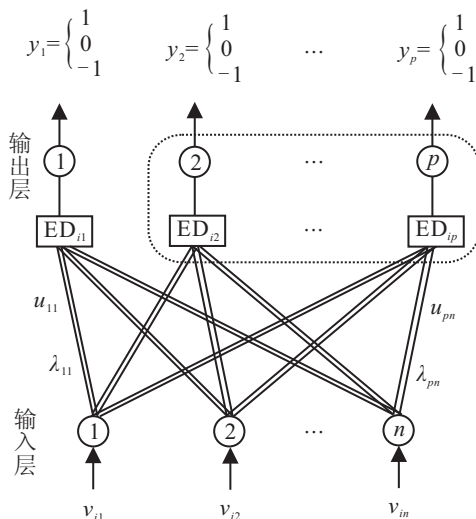


图2 改进ENN2聚类算法模型结构

值输出, 即样本  $M_i$  存在确定属于故障类型  $T_t$  (确定性故障), 可能属于故障类型  $T_t$  (潜在故障) 和确定不属于故障类型  $T_t$  三种状态. 综上所述, 改进ENN2聚类算法模型结构如图2所示.

2.2 学习算法

ENN2和改进ENN2聚类算法均采用无导师监督学习算法, 在执行过程中采用跟随领导策略, 既不需要初始化聚类个数, 也不需要初始化聚类中心. 对学习算法的改进包括修改聚类中心更新公式和采用新的聚类策略, 改进后学习算法的具体步骤如下.

Step 1: 参数初始化. 初始化样本集特征  $c_j$  的经典域  $V_j = \langle v_j^C, v_j^D \rangle$  和可变距离参数  $\lambda_{j0}$ , 其中

$$v_j^C = \min_{i=1,2,\dots,m} (v_{ij}),$$

$$v_j^D = \max_{i=1,2,\dots,m} (v_{ij}),$$

$$\lambda_{j0} = \frac{v_j^D - v_j^C}{\sqrt{m}}, \quad j = 1, 2, \dots, n,$$

$m$  为样本集的元素个数.

Step 2: 创建故障类型  $T_1$ . 令  $i = 1$ , 取样本  $M_1$ , 创建故障类型  $T_1 = \{M_1\}$ , 此时故障类型  $T_1$  的元素个数  $N_1 = 1$ , 聚类中心

$$z_1 = \left\{ u_{1j} = \frac{v_{1j}}{N_1}, j = 1, 2, \dots, n \right\}.$$

特征  $c_j$  的经典域范围为  $\langle v_{1j}^A, v_{1j}^B \rangle$ , 其中

$$v_{1j}^A = u_{1j} - \lambda_{0j}, \quad v_{1j}^B = u_{1j} + \lambda_{0j},$$

$$j = 1, 2, \dots, n.$$

Step 3: 取样本  $M_i$ . 令  $i = i + 1$ , 取样本  $M_i$ , 假设此时已有  $p$  种故障类型  $T_1, T_2, \dots, T_p$ , 且故障类型  $T_t$  的聚类中心为  $z_t$ , 包含样本个数为  $N_t$ , 对应特征  $c_j$  的经典域为  $\langle v_{tj}^A, v_{tj}^B \rangle, j = 1, 2, \dots, n, t = 1, 2, \dots, p$ .

Step 4: 确定样本  $M_i$  的故障类型. 根据式(1)计算样本  $M_i$  与故障类型  $T_t$  的可拓距离  $ED_{it}$ . 如果

$$ED_{it} \geq 0, \quad t = 1, 2, \dots, p,$$

则表示样本  $M_i$  属于故障类型  $T_t$ , 此时

$$T_t = T_t \cup \{M_i\}, \quad N_t = N_t + 1,$$

更新  $T_t$  的聚类中心

$$z'_t = \left\{ u'_{tj} = \frac{v_{tj} + (N_t - 1)u_{tj}}{N_t}, j = 1, 2, \dots, n \right\} \quad (2)$$

和可变距离参数

$$\lambda'_{tj} = \max\{\lambda_{0j}, \max_{M_{s_1}, M_{s_2} \in T_t} (|v_{s_1j} - v_{s_2j}|)\}, \quad (3)$$

其中  $\max_{M_{s_1}, M_{s_2} \in T_t} (|v_{s_1j} - v_{s_2j}|)$  为故障类型属于  $T_t$  的样本中特征  $c_j$  的最大距离. 如果

$$0 > ED_{it} \geq \delta > -1,$$

则表示样本  $M_i$  可能属于故障类型  $T_t$ , 此时  $T_t = T_t \cup \{M_i\}$ , 但不更新样本个数  $N_t$ 、聚类中心  $z_t$  和特征  $c_j$  的经典域  $V_{0tj}$ . 如果对于当前所有故障类型  $T_t$  ( $t = 1, 2, \dots, p$ ), 均有  $\delta > ED_{it}$ , 则表示样本  $M_i$  不属于已有的故障类型, 此时创建新的故障类型  $T_{p+1} = \{M_i\}$ , 包含元素个数  $N_{p+1} = 1$ , 聚类中心为

$$z_{p+1} = \left\{ u_{(p+1)j} = \frac{v_{(p+1)j}}{N_{p+1}}, j = 1, 2, \dots, n \right\},$$

对应特征  $c_j$  的经典域范围为  $\langle v_{(p+1)j}^A, v_{(p+1)j}^B \rangle$ .

**Step 5:** 聚类中心更新后, 重新聚类所有诊断为可能故障的样本. 当故障类型  $T_t$  的聚类中心发生更新时, 对于任意一个进行过聚类的样本  $M_i$ , 如果  $0 > ED_{it} \geq \delta > -1$ , 则需要重新判断样本  $M_i$  是否属于更新后的故障类型  $T_t$ .

**Step 6:** 新故障类型产生时, 判断所有诊断过的样本是否属于新故障. 当产生新故障类型  $T_{t+1}$  时, 对于任意一个进行过聚类的样本  $M_i$ , 都需要判断样本  $M_i$  是否属于新故障类型  $T_{p+1}$ .

**Step 7:** 终止条件. 重复 Step 3 ~ Step 6, 直到  $i = m$ .

### 3 算法性能分析

从模型参数和时间复杂度两个方面对改进 ENN2 聚类算法的性能进行分析.

#### 3.1 参数分析

在改进 ENN2 聚类算法中, 因为阈值  $\delta$  的选取具有主观性, 所以主要讨论阈值  $\delta$  对诊断结果的影响. 通过计算样本  $M_i$  与聚类中心  $z_t$  的可拓距离  $ED_{it}$  判断样本  $M_i$  是否属于故障类型  $T_t$ . 当  $ED_{it} \geq 0$  时, 样本  $M_i$  属于故障类型  $T_t$ , 即为确定性故障; 当  $0 > ED_{it} \geq \delta > -1$  时, 样本  $M_i$  可能属于故障类型  $T_t$ , 即为潜在故障; 当  $\delta > ED_{it}$  时, 样本  $M_i$  不属于故障类型  $T_t$ . 故障类型可能是  $T_t$  的样本越多, 更新聚类中心  $z_t$  后需要重新诊断的样本越多. 当  $\delta \rightarrow -\infty$  时, 没有故障类型不是  $T_t$  的样本, 当  $\delta \rightarrow 0$  时, 没有故障类型可能是  $T_t$  的样本, 因此阈值  $\delta$  的选取不会影响故障类型确定是  $T_t$  的样本的诊断结果, 但会影响故障类型可能是和确定不是  $T_t$  的样本的诊断结果, 通常取  $\delta = -0.5$ .

#### 3.2 时间复杂度分析

ENN2 和改进 ENN2 聚类算法的主要时间代价是计算样本与各聚类中心的可拓距离和聚类中心的更新. 假设每个样本仅具有一个属性, 已有  $m$  个样本进行过聚类, 得到  $p$  种故障, 比较第  $m+1$  个样本聚类时两种算法的时间复杂度.

对于 ENN2, 首先, 需要计算第  $m+1$  个样本和已有  $p$  个聚类中心的可拓距离, 计算  $p$  次; 然后, 对已经

聚类过的  $m$  个样本重新聚类, 计算  $m \times p$  次; 最后, 如果已有  $m$  个样本的故障类型都发生变化, 则最多需要更新聚类中心的上限和下限共  $2m$  次, 所以每增加一个样本, ENN2 的时间复杂度为  $O(p + m(p+2))$ . 对于改进 ENN2 聚类算法, 首先同样计算第  $m+1$  个样本与已有  $p$  个聚类中心的可拓距离, 计算  $p$  次; 如果产生新的聚类, 则需要判断已有的  $m$  个样本是否属于新的聚类, 计算  $m$  次; 如果没有新的聚类, 则需要对诊断为可能发生故障的共  $m'$  个样本进行重新聚类, 计算  $m'p$  次. 对诊断为确定性故障的样本不需要重新聚类, 所以改进 ENN2 聚类算法的时间复杂度为  $O(p + \max(m, m'p))$ . 因为诊断为可能发生故障的样本个数少于样本总个数, 即  $m' < m$ , 所以  $\max(m, m'p) < m(p+2)$ , 即改进 ENN2 聚类算法的时间复杂度低于 ENN2 的时间复杂度.

下面证明聚类中心更新时, 诊断为确定性故障的样本故障类型不会发生变化. 原命题等价于对于任意  $M_s \in T_t$ ,  $v_{tj}^B \geq v_{sj} \geq v_{tj}^A$ , 当聚类中心  $z_t$  更新为  $z'_t$  时, 仍有  $M_s \in T'$ ,  $v_{tj}^{B'} \geq v_{sj} \geq v_{tj}^{A'}$ .

由式 (2) 可知, 故障类型  $T_t$  的聚类中心由  $z_t$  更新为  $z'_t$  时, 聚类中心的更新公式为

$$u'_{tj} = \frac{v_{ij} + (N_t - 1)u_{tj}}{N_t},$$

其中  $v_{ij}$  为新加入故障类型  $T_t$  的样本  $M_i$  的第  $j$  个特征值.

当  $v_{ij}$  在聚类中心左侧, 即  $u_{tj} \geq v_{ij} \geq v_{tj}^A$  时, 有

$$u_{tj} \geq u'_{tj} = \frac{v_{ij} + (N_t - 1)u_{tj}}{N_t} \geq v_{tj}^A,$$

所以有

$$v_{tj}^{A'} + \lambda'_{tj} \geq v_{tj}^A.$$

又因为

$$v_{tj}^A - v_{tj}^{A'} = u_{tj} - u'_{tj},$$

所以有

$$\lambda'_{tj} \geq v_{tj}^A - v_{tj}^{A'} = u_{tj} - u'_{tj},$$

即

$$u_{tj} \leq u'_{tj} + \lambda' = v_{tj}^{B'}. \quad (4)$$

由式 (3) 可知,  $\lambda'_{tj} \geq \lambda_{tj}$ , 所以

$$u'_{tj} - \lambda'_{tj} \leq u_{tj} - \lambda_{tj},$$

即

$$v_{sj} \geq v_{tj}^A \geq v_{tj}^{A'}. \quad (5)$$

综合式 (4) 和 (5) 可知, 当新增样本  $M_i$  的特征值满足  $u_{tj} \geq v_{ij} \geq v_{tj}^A$  时, 对于任意  $M_s \in T_t$ , 其特征值都满足  $v_{tj}^{B'} \geq v_{sj} \geq v_{tj}^{A'}$ . 同理, 当  $v_{ij}$  在聚类中心右侧, 即  $v_{tj}^B \geq v_{ij} \geq u_{tj}$  时, 对于任意  $M_s \in T_t$ , 其特征值都满足  $v_{tj}^{B'} \geq v_{sj} \geq v_{tj}^{A'}$ .

综上所述, 对于任意  $M_s \in T_t, v_{tj}^B \geq v_{sj} \geq v_{tj}^A$ , 当聚类中心  $z_t$  更新为  $z_t'$  时, 仍有  $M_s \in T', v_{tj}^{B'} \geq v_{sj} \geq v_{tj}^A$ .

### 4 实例验证

采用文献[14]中24组汽轮发电机组振动信号的频谱数据进行实例验证, 每组数据包含5个特征, 分别为小于  $0.4f$ 、 $(0.4 \sim 0.5)f$ 、 $1f$ 、 $2f$  和大于等于  $3f$  ( $f$  为一阶转频). 在文献[14]中, 前15组数据用于训练, 后9组数据用于测试, 而本文所提出的方法不需要训练, 将前15组数据用于离线诊断, 后9组数据用于在线诊断, 验证其多故障诊断和发现潜在故障的能力.

采用改进ENN2聚类算法, 对前15组数据进行离线诊断, 得到5种故障类型, 每个样本和每种故障类型聚类中心的可拓距离见表1.

表1 离线诊断结果

序号	$T_1$	$T_2$	$T_3$	$T_4$	$T_5$
1	<b>0.013</b>	-0.745	-0.948	-0.986	-0.995
2	-1.000	<b>0.000</b>	-1.000	-1.000	-1.000
3	<b>-0.421</b>	<b>-0.162</b>	-0.927	-0.977	-0.984
4	<b>0.006</b>	<b>-0.417</b>	-0.938	-0.997	-0.998
5	<b>0.000</b>	-0.581	-1.000	-1.000	-1.000
6	-0.965	-0.958	<b>0.060</b>	-0.953	-0.936
7	-1.000	-1.000	<b>0.000</b>	-1.000	-1.000
8	-0.929	-1.000	<b>0.000</b>	-0.705	-0.715
9	-0.971	-0.979	<b>0.056</b>	-0.818	-0.883
10	-0.985	-0.989	<b>0.030</b>	-0.899	-0.967
11	-0.961	-0.972	-0.635	<b>0.077</b>	<b>0.025</b>
12	-0.988	-0.992	-0.578	<b>0.023</b>	-0.576
13	-0.958	-0.988	-0.513	<b>0.029</b>	<b>-0.290</b>
14	-1.000	-1.000	-0.791	<b>0.000</b>	<b>-0.440</b>
15	-1.000	-1.000	-1.000	-1.000	<b>0.000</b>

取阈值  $\delta = -0.5$ , 表1中加黑显示的部分表示样本与其对应故障类型的可拓距离, 如样本  $M_1$  的故障类型确定是  $T_1$ ; 样本  $M_3$  的故障类型可能是  $T_1$  和  $T_2$ ; 样本  $M_4$  的确定性故障类型是  $T_1$ , 同时可能故障类型是  $T_2$ ; 样本  $M_{11}$  的故障类型确定是  $T_4$  和  $T_5$ .

由于已知前5组样本数据对应的故障类型是油膜振动, 中间5组数据对应的故障类型是不平衡, 后5组数据对应的故障类型是不对中, 可以对初始诊断结果进行改进, 合并故障类型  $T_1$ 、 $T_2$  和故障类型  $T_4$ 、 $T_5$ , 得到故障类型  $T'_1$  和  $T'_3$ .

合并聚类中心的公式为

$$z_t = \left\{ u_{tj} = \frac{N_{t_1j}u_{t_1j} + N_{t_2j}u_{t_2j}}{N_{t_1j} + N_{t_2j}}, j = 1, 2, \dots, n \right\},$$

其中  $u_{t_1j}$ 、 $u_{t_2j}$  和  $N_{t_1j}$ 、 $N_{t_2j}$  分别为故障类型  $T_{t_1}$  和故障类型  $T_{t_2}$  的聚类中心和所包含的样本个数. 合并可

变距离参数的公式为

$$\lambda_{tj} = \max\{\lambda_{t_1j}, \lambda_{t_2j}, \max_{M_{s_1} \in T_{t_1}, M_{s_2} \in T_{t_2}} (|v_{s_1j} - v_{s_2j}|)\}.$$

其中:  $\lambda_{t_1j}$ 、 $\lambda_{t_2j}$  分别为故障类型  $T_{t_1}$  和故障类型  $T_{t_2}$  的可变距离参数,  $\max_{M_{s_1} \in T_{t_1}, M_{s_2} \in T_{t_2}} (|v_{s_1j} - v_{s_2j}|)$  为合并的所有样本中特征  $c_j$  的最大距离. 合并后, 得到15个样本对应3种故障类型的可拓距离见表2. 表2中, 加黑显示的部分表示样本与其对应故障类型的可拓距离, 诊断结果与实际情况一致, 所以改进ENN2聚类算法具有较好的离线诊断能力.

表2 合并故障类型后诊断结果

序号	$T'_1$	$T_2$	$T'_3$
1	<b>0.012</b>	-0.948	-0.965
2	<b>0.000</b>	-1.000	-1.000
3	<b>0.039</b>	-0.927	-0.974
4	<b>0.006</b>	-0.938	-0.996
5	<b>0.000</b>	-1.000	-1.000
6	-0.960	<b>0.060</b>	-0.941
7	-1.000	<b>0.000</b>	-1.000
8	-0.930	<b>0.000</b>	-0.631
9	-0.972	<b>0.056</b>	-0.773
10	-0.985	<b>0.030</b>	-0.758
11	-0.962	-0.635	<b>0.072</b>
12	-0.989	-0.578	<b>0.022</b>
13	-0.959	-0.513	<b>0.029</b>
14	-1.000	-0.791	<b>0.000</b>
15	-1.000	-1.000	<b>0.000</b>

采用得到的各种故障的聚类中心和可变距离参数直接对后9组测试数据进行在线诊断, 诊断结果见表3. 表3中, 加黑显示的部分表示样本与其对应故障类型的可拓距离, 其中对于样本  $M_{16}$  和  $M_{17}$ , 本文的诊断结果为可能存在故障  $T'_1$ , 文献[14]的诊断结果为故障  $T'_1$ , 分析表明, 样本  $M_{16}$  和  $M_{17}$  可能处在故障  $T'_1$  的早期阶段; 对于样本  $M_{21}$ , 本文的诊断结果为确定存在故障  $T_2$ , 可能存在故障  $T'_3$ , 文献[14]的诊断结果为故障  $T_2$ , 分析表明, 由于文献[14]方法不具备多故障诊断和潜在故障诊断的能力, 可能漏报故障  $T'_3$ ; 其他样本的诊断结果与文献[14]一致.

表3 在线诊断结果

序号	$T'_1$	$T_2$	$T'_3$
16	<b>-0.216</b>	-1.319	-1.315
17	<b>-0.061</b>	-1.164	-1.312
18	<b>0.033</b>	-0.978	-0.941
19	-0.958	<b>0.027</b>	-0.938
20	-0.868	<b>0.130</b>	-0.572
21	-0.936	<b>0.042</b>	<b>-0.180</b>
22	-0.951	-0.688	<b>0.094</b>
23	-0.774	-0.619	<b>0.186</b>
24	-0.983	-0.754	<b>0.032</b>

综上所述,基于改进 ENN2 聚类算法的多故障诊断方法不仅具有多故障诊断能力(如在离线诊断初期可以诊断出样本  $M_{11}$  的多故障类型是  $T_4$  和  $T_5$ ),而且具有发现潜在故障的能力(如在在线诊断阶段可以发现样本  $M_{21}$  可能具有潜在故障  $T_3$ )。相比之下,未改进的 ENN2 只能进行单故障诊断,不具备多故障诊断和发现潜在故障的能力。

## 5 结 论

本文首先建立了多故障诊断问题模型,并提出利用聚类算法对其进行求解;然后针对 ENN2 的不足,以及其在多故障诊断应用中的缺陷,提出了基于改进 ENN2 聚类算法的多故障诊断方法,并对其参数和时间复杂度进行了分析,证明了该方法的时间复杂度低于 ENN2 的时间复杂度;最后采用汽轮发电机组振动信号频谱数据对方法进行验证,用前 15 组数据进行离线的多故障诊断,并将得到的诊断模型用于后 9 组数据的在线状态监控。结果表明,所提出的多故障诊断方法不仅能解决多故障诊断问题,而且为潜在故障的早期诊断提供了手段。

## 参考文献(References)

- [1] Jing Jianping, Meng Guang. A novel method for multi-fault diagnosis of rotor system[J]. Mechanism and Machine Theory, 2009, 44(4): 697-709.
- [2] Saimurugan M, Ramachandran K J, Sugumaran V, et al. Multi component fault diagnosis of rotational mechanical system based on decision tree and support vector machine[J]. Expert System with Applications, 2011, 38(4): 3819-3826.
- [3] Wang M H. Extension neural network for power transformer incipient fault diagnosis[J]. IEEE Proc Gener Trans on Distrib, 2003, 150(6): 679-685.
- [4] 周玉, 钱旭, 张俊彩, 等. 可拓神经网络研究综述[J]. 计算机应用研究, 2010, 27(1): 1-5.  
(Zhou Y, Qian X, Zhang J C, et al. Survey and research of extension neural network[J]. Application Research of Computers, 2010, 27(1): 1-5.)
- [5] Wang M H, Hung C P. Extension neural network[C]. Proc of the Int Joint Conf on Neural Networks. Oregon, 2003: 339-403.
- [6] Wang M H, Hung C P. Extension neural network and its applications[J]. IEEE Trans on Neural Networks, 2003, 16(5): 779-784.
- [7] Wang M H, Hung C P. Extension neural network-type 2 and its applications[J]. IEEE Trans on Neural Networks, 2005, 16(6): 1352-1361.
- [8] Ye J. Application of extension theory in misfire fault diagnosis of gasoline engines[J]. Expert Systems with Application, 2009, 36(2): 1217-1221.
- [9] Kuei-Hsiang Chao, Chia-Lung Chiu, Ching-Ju Li, et al. A novel neural network with simple learning algorithm for islanding phenomenon detection of photovoltaic systems[J]. Expert Systems with Applications, 2011, 38(10): 12107-12115.
- [10] Hung-cheng Chen, Feng-chang Gu, Meng-hui Wang. A novel extension neural network based partial discharge pattern recognition method for high-voltage power apparatus[J]. Expert Systems With Applications, 2012, 39(3): 3423-3431.
- [11] Mohamed S, Tettey T, Marwala T. An extension neural network and genetic algorithm for bearing fault classification[C]. Proc of Int Joint Conf on Neural Networks. Vancouver, 2006: 3941-3948.
- [12] 徐圆, 冯晶, 朱群雄. 基于可拓理论的 RBF 神经网络研究及应用[J]. 控制与决策, 2011, 26(11): 1721-1725.  
(Xu Y, Feng J, Zhu Q X. Research and application of extension theory-based radial basis function neural network[J]. Control and Decision, 2011, 26(11): 1721-1725.)
- [13] Yang Chunyan, Cai Wen. Extensics theory, method and application[M]. Beijing: Science Press, 2013: 87-100.
- [14] 王永林. 汽轮发电机的粒子群可拓故障诊断算法[J]. 电力系统保护与控制, 2010, 38(17): 87-90.  
(Wang Y L. Extension fault diagnosis for turbine-generator set based on PSO algorithm[J]. Power System Protection and Control, 2010, 38(17): 87-90.)

(责任编辑: 郑晓蕾)