http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2017.0632

基于近邻传播聚类的航空电子部件 LMK 诊断模型

张伟[1], 许爱强[1,*], 平殿发[1], 夏菲[2]
(1. 海军航空大学, 烟台 264001; 2. 国网辽阳供电公司 信息通信分公司, 辽阳 111000)

摘 要: 针对小样本条件下,航空电子部件功能模块故障诊断精度不高的问题,将局部多核学习(LMKL)算法的多分辨率解释与局部特征自适应表示能力和超限学习机(ELM)运算高效的特点相结合,提出一种新的局部聚类 MK-ELM(LCMKELM)诊断模型。通过引入近邻传播(AP)聚类,在挖掘训练样本局部特征信息的同时,有效约减了局部算法的计算复杂性,避免了过学习问题的出现;通过分别分析输入空间与特征空间的聚类特征,构造了相应的2种选通函数 M_1、M_2,以优化选通函数的模型参数取代优化局部权重,有效解决了核超限学习机(KELM)的对偶优化形式关于局部权重二次非凸的问题。将本文模型应用于某型机旋转变压器激励发生电路功能模块故障诊断,结果表明:相比于4种常用的多核诊断算法,模型在实现低漏警、低虚警的同时,采用 M_1 选通函数的诊断算法将诊断精度平均提升了3.80%,采用 M_2 选通函数的诊断算法将诊断精度平均提升了5.98%。同时,模型在实现与流行的 LMKL 算法相近的训练时间的同时,测试时间更短。

关 键 词: 故障诊断; 多核学习; 交替优化; 近邻传播(AP); 选通模型; 局部算法
中图分类号: V243; TP181
文献标识码: A **文章编号**: 1001-5965(2018)08-1693-12

针对航空电子设备故障特点开展故障诊断算法及其应用的研究,已成为提升航空保障能力的关键。航空电子设备及其子系统故障模式多样[1-4],各功能模块交联关系复杂,这些存在的问题使得征兆与故障之间不具有严格的对应关系,增加了故障诊断的难度和复杂性[5-7]。

当前,基于核的学习算法,例如支持向量机(Support Vector Machine, SVM)、核超限学习机(Kernel Extreme Learning Machine, KELM)等,已经被广泛应用到故障诊断领域中,并且展示出优越的性能[8-9]。多核学习(Multi-Kernel Learning, MKL)作为一种灵活性更强、解释能力更好的核学习算法[10],近些年来受到了广泛的关注。文献[11-12]已经将 MKL 引入到了包括模拟电路、变压器等系统的故障诊断中,并通过大量实例证明基于 MKL 的诊断算法在面对单故障、多故障以及并发故障时均能取得更高的诊断精度。

目前,大多数的文献主要面向基于 SVM 的 MKL 算法研究,而对于 MKL 的具体形式则主要采用有限个给定基核的线性凸组合。半定规划,二次约束二次规划,半无限线性规划等算法相继被提出去解 MKL 问题。文献[13]在2008年提出了一种更有效的算法,称作 SimpleMKL,其使得 MKL 对于大样本问题更加实用。但在这些算法中,每个基核的权重在整个输入空间保持不变,显然忽视了输入空间的局部多样性。不同的核函数

收稿日期: 2017-10-16; **录用日期**: 2017-11-17; **网络出版时间**: 2018-01-25 11:04
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20180123.1858.003.html
基金项目: 国家自然科学基金(61571454); 山东省自然科学基金(ZR2016FQ03)
* **通信作者**. E-mail: hjhyautotest@sina.com

引用格式: 张伟, 许爱强, 平殿发, 等. 基于近邻传播聚类的航空电子部件 LMK 诊断模型[J]. 北京航空航天大学学报, 2018, 44(8): 1693-1704. ZHANG W, XU A Q, PING D F, et al. Localized multi-kernel diagnosis model for avionics based on affinity propagation clustering[J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(8): 1693-1704 (in Chinese).

针对不同的样本具有不同的可用性,为了体现这种可用性上的差异性,有必要为每一个样本点调节核系数[14-15]。基于这种思想,局部算法被引入 MKL 中,即采用一种依赖样本的方式为基核分配权重。为此,科研人员已经开展了相关研究。

Gönen 和 Alpaydin 提出了最初的局部多核学习(Localized Multi-kernel Learning, LMKL)算法[16]。取代直接优化局部权重,文献[16]采用一系列选通函数去近似局部权重,以解决局部权重的优化形式二次非凸的问题。文献[17-18]基于文献[16]的思想,提出 2 种改进算法。前者依据半径-间距界将最小超球的半径融入到 LMKL 中;后者构造了一种概率置信核以此充当选通函数。文献[19]提出了一个 p 范数约束的 LMKL 模型。通过使用二阶 Taylor 展开来近似 p 范数,将局部权重的优化问题转化为一个非凸的单约束二次规划问题,最终采用半定规划进行求解。LMKL 算法将局部权重拟合到每一个确定样本上,虽然可以充分描述样本的类内多样性,但也可能会导致过拟合问题的出现。同时,需要优化的变量个数等于样本数乘以基核数,当数据规模较大时算法的复杂性将变得无法控制。

为此,考虑到来自于同一聚类的数据具有相似的局部特征,趋向于具有相似的性质。本文结合航空电子设备及其子系统故障诊断的实际,提出一种基于近邻传播(Affinity Propagation, AP)局部聚类 MK-ELM(LCMKELM)诊断模型。该模型通过为具有相似分布特性的故障样本分配相同的局部权重,在继承一般 LMKL 算法局部特征自适应表示能力的同时,避免了过学习风险的出现,有效约减了局部算法的计算复杂性。同时,通过分别构造面向输入空间和特征空间的选通函数,并以选通函数近似局部权重,提出一种 2 步交替优化的策略,有效解决了局部权重二次非凸的优化形式难以求解的问题。

1 问题描述

对于一个具有 c($c\geqslant 2$)种故障模式的诊断问题,设训练和测试数据集分别为 $\boldsymbol{D}_{\mathrm{Tr}}$ 和 $\boldsymbol{D}_{\mathrm{Te}}$。令 $\boldsymbol{D}_{\mathrm{Tr}}=\{(\boldsymbol{x}_1,y_1),(\boldsymbol{x}_2,y_2),\cdots,(\boldsymbol{x}_i,y_i),\cdots,(\boldsymbol{x}_n,y_n)\}$。其中:$\boldsymbol{x}_i\in\mathbf{R}^d$ 为训练数据实例,$d\in\mathbf{R}$ 为 $\boldsymbol{x}_i$ 的维数;$y_i\in\{1,2,\cdots,c\}$ 为 $\boldsymbol{x}_i$ 对应的故障模式。令 $\boldsymbol{D}_{\mathrm{Te}}=\{(\boldsymbol{x}'_1,y'_1),(\boldsymbol{x}'_2,y'_2),\cdots,(\boldsymbol{x}'_j,y'_j),\cdots,(\boldsymbol{x}'_{n'},y'_{n'})\}$。其中,$\boldsymbol{x}'_j\in\mathbf{R}^d$ 为测试数据实例;y'_j 未知。故障诊断模型设计的目的就是要基于 $\boldsymbol{D}_{\mathrm{Tr}}$ 寻找一个映射函数 $f(\cdot):\mathbf{R}^d\rightarrow\mathbf{R}$,以使得 $\boldsymbol{D}_{\mathrm{Te}}$ 中的任一个数据实例 $\boldsymbol{x}'_j$ 都可以通过映射得到相应的故障模式 $y'_j\in\{1,2,\cdots,c\}$。

LMKL 中局部权重定义为 $\boldsymbol{\gamma}=\{\gamma_q(\boldsymbol{x}_i)\}_{q=1,2,\cdots,r}^{i=1,2,\cdots,n}$,每一个局部权重既与基核相关,又与对应的样本相关。设 $\{k_1(\cdot,\cdot),k_2(\cdot,\cdot),\cdots,k_q(\cdot,\cdot),\cdots,k_r(\cdot,\cdot)\}$ 为预定义的 r 个基核;$\{\phi_1(\cdot),\phi_2(\cdot),\cdots,\phi_q(\cdot),\cdots,\phi_r(\cdot)\}$ 为对应于基核的特征映射;$\{\boldsymbol{K}_1,\boldsymbol{K}_2,\cdots,\boldsymbol{K}_q,\cdots,\boldsymbol{K}_r\}$ 为数据集通过基核得到的核矩阵。则根据 ELM 理论,LMK-ELM 的初始优化问题定义为

$$\begin{aligned}&\min_{\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\xi}}\left(\frac{1}{2}\sum_{q=1}^{r}\|\boldsymbol{\beta}_q\|_{\mathrm{F}}^2+\frac{C}{2}\sum_{i=1}^{n}\|\boldsymbol{\xi}_i\|_2^2\right)\\&\text{s.t.}\sum_{q=1}^{r}\gamma_q(\boldsymbol{x}_i)\phi_q(\boldsymbol{x}_i)\boldsymbol{\beta}_q=\boldsymbol{Y}_i^{\mathrm{T}}-\boldsymbol{\xi}_i^{\mathrm{T}}\\&\quad i=1,2,\cdots,n\end{aligned}\tag{1}$$

式中:$\boldsymbol{\beta}=[\boldsymbol{\beta}_1^{\mathrm{T}},\boldsymbol{\beta}_2^{\mathrm{T}},\cdots,\boldsymbol{\beta}_r^{\mathrm{T}}]$ 为 ELM 的输出总权重;$\boldsymbol{\beta}_q\in\mathbf{R}^{|\phi_q(\cdot)|\times m}$ 为基于第 q 个基核的 ELM 输出权重,$|\phi_q(\cdot)|$ 为第 q 个特征空间的维数,m 为 ELM 输出节点个数;$\boldsymbol{\xi}_i=[\xi_{i1},\xi_{i2},\cdots,\xi_{im}]^{\mathrm{T}}$ 为模型对应于训练样本 $\boldsymbol{x}_i$ 的训练误差向量,且有 $\boldsymbol{\xi}=[\boldsymbol{\xi}_1,\boldsymbol{\xi}_2,\cdots,\boldsymbol{\xi}_n]^{\mathrm{T}}$;$\boldsymbol{Y}_i=[Y_{i1},Y_{i2},\cdots,Y_{im}]^{\mathrm{T}}$ 为对应于训练样本 $\boldsymbol{x}_i$ 的理想输出向量,且有 $\boldsymbol{Y}=[\boldsymbol{Y}_1,\boldsymbol{Y}_2,\cdots,\boldsymbol{Y}_n]^{\mathrm{T}}$;$C$ 为正则化因子。

基于初始优化问题的 Lagrange 函数,可以进一步得到 LMK-ELM 的对偶优化问题:

$$\begin{aligned}&\max_{\boldsymbol{\alpha}}\min_{\boldsymbol{\gamma}}\left(-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\boldsymbol{\alpha}_i^{\mathrm{T}}\boldsymbol{\alpha}_j\sum_{q=1}^{r}\gamma_q(\boldsymbol{x}_i)\gamma_q(\boldsymbol{x}_j)k_q(\boldsymbol{x}_i,\boldsymbol{x}_j)-\right.\\&\left.\frac{1}{2C}\sum_{i=1}^{n}\boldsymbol{\alpha}_i^{\mathrm{T}}\boldsymbol{\alpha}_i+\sum_{i=1}^{n}\boldsymbol{\alpha}_i^{\mathrm{T}}\boldsymbol{Y}_i\right)\end{aligned}\tag{2}$$

式中:$\boldsymbol{\alpha}_i$ 为 Lagrange 乘子,对应于 ELM 的模型参数,且 $\boldsymbol{\alpha}=[\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n]^{\mathrm{T}}$,$\boldsymbol{\alpha}_i=[\alpha_{i1},\alpha_{i2},\cdots,\alpha_{im}]^{\mathrm{T}}$。

为了解上述的对偶优化问题,通常采用一种交替优化的策略交替更新 ELM 模型参数和局部核权重,最终得到决策函数:

$$f(\boldsymbol{x})=\sum_{i=1}^{n}\boldsymbol{\alpha}_i^{\mathrm{T}}\sum_{q=1}^{r}\gamma_q(\boldsymbol{x})\gamma_q(\boldsymbol{x}_i)k_q(\boldsymbol{x},\boldsymbol{x}_i)\tag{3}$$

无论怎样,在优化过程中:①局部权重的数量等于基核数乘以样本数,算法复杂性较高,且存在过学习风险;②在式(2)中关于局部权重的优化是一个困难的二次非凸问题。

针对上述问题,本文取代一般 LMKL 算法中将局部权重拟合到每一个确定样本上,而是将局部权重拟合到空间中具有相似局部结构的一类群组上,进而提出一种新的面向聚类的 LMK-ELM 模型。该模型的基本框架如图 1 所示。

http://bhxb.buaa.edu.cn

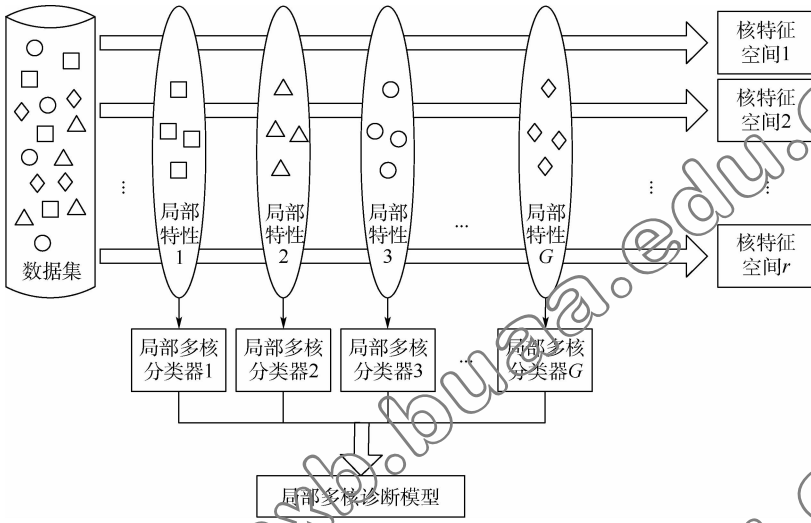


图 1　本文模型的基本框架

Fig. 1　General framework of proposed model

2　LCMKELM 诊断模型

2.1　基于 AP 算法的样本空间划分

为挖掘故障样本中隐含的局部分布特征,本节引入 AP 算法[20-21]。相比于经典的 K-均值聚类,AP 算法不需要指定聚类的数目,对初始值的选取不敏感,并且文献[20]证明其具有更低的误差平方和。这使得在 AP 算法中先验经验成为应用的非必要条件,因而更适合本文的应用背景。

首先,定义样本间的相似性度量准则为

$$\varphi(i,j) = -\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2 \tag{4}$$

显然,$\varphi(i,j)$越大,则 2 个数据点之间的相似性越高。进一步得到 n 个样本点组成的相似度矩阵 $\boldsymbol{S}$,即

$$\boldsymbol{S} = \begin{bmatrix} \varphi(1,1) & \cdots & \varphi(1,n) \\ \vdots & & \vdots \\ \varphi(n,1) & \cdots & \varphi(n,n) \end{bmatrix}$$

AP 算法中传递 2 种类型的消息,分别称作吸引信息和归属信息,并存储在 2 个矩阵 $\boldsymbol{V}$ 和 $\boldsymbol{W}$ 中。$\boldsymbol{V}$ 中的 $v(i,j)$描述了数据对象 j 适合作为数据对象 i 的聚类中心的程度,表示的是从 i 到 j 的消息。而 $\boldsymbol{W}$中的 $w(i,j)$描述了数据对象 i 选择数据对象 j 作为其聚类中心的适合程度,表示的是从 j 到 i 的消息。$v(i,j)$与 $w(i,j)$越强,则 j 点作为聚类中心的可能性就越大,并且 i 点隶属于以 j 点为聚类中心的聚类的可能性也越大。吸引信息和归属信息的更新方程分别定义为

$$v_{t+1}(i,j) = \varphi(i,j) - \max_{u \neq j}\{w_t(i,u) + \varphi(i,u)\}$$

$$w_{t+1}(i,j) = \begin{cases} \min\left\{0, v_t(j,j) + \sum_{u \neq \{i,j\}} \max\{0, v_t(u,j)\}\right\} & i \neq j \\ \sum_{u \neq j} \max\{0, v_t(u,j)\} & i = j \end{cases}$$

式中:下标 t 为当前迭代次数。

为了避免振荡,AP 算法更新信息时引入了衰减系数 $\lambda \in (0,1)$。通常,λ 越接近于 0,振荡越大,收敛越慢,生成的聚类个数较多;而 λ 越接近于 1,振荡越小,收敛越快,生成的聚类个数也相对较少。为此,进一步令

$$v_{t+1}(i,j) \leftarrow (1-\lambda)v_{t+1}(i,j) + \lambda v_t(i,j)$$

$$w_{t+1}(i,j) \leftarrow (1-\lambda)w_{t+1}(i,j) + \lambda w_t(i,j)$$

$\boldsymbol{V}$ 和 $\boldsymbol{W}$ 初始均为全 0 矩阵,AP 算法通过迭代过程不断更新每一个点的吸引度和归属度,如果决策信息经过若干次迭代之后保持不变或者算法执行超过设定的最大迭代次数,则算法终止。此时,决策矩阵定义为 $\boldsymbol{V}+\boldsymbol{W}$,决策矩阵主对角线元素$\{v(i,i)+w(i,i)\}$中所有大于 0 的 G 个点被选择作为聚类中心,其余的点根据式(5)定义的规则被分配到相应的聚类中:

$$c(\boldsymbol{x}_i) = \arg\max_{g \in \{1,2,\cdots,G\}} (v(i,o_g) + w(i,o_g)) \tag{5}$$

式中:o_g 为第 g 个聚类中心在原始训练数据集中对应的样本序号;$c(\boldsymbol{x}_i)$为 $\boldsymbol{x}_i$ 所对应的聚类标签。

这样,数据实例集合 $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n\}$将被划分为 G 个子集,即 $\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_G$,并且每个子集满足:①$\boldsymbol{X}_s \neq \varnothing, s = 1,2,\cdots,G$;②$\bigcup_{s=1}^{G} \boldsymbol{X}_s = \boldsymbol{X}$;③$\boldsymbol{X}_s \cap \boldsymbol{X}_{s'} = \varnothing, s, s' = 1,2,\cdots,G$,且 $s \neq s'$。

http://bhxb.buaa.edu.cn

2.2 LCMKELM 诊断模型优化

根据 2.1 节的聚类结果,当将局部权重拟合到每一个聚类上时,新的局部权重可以表示为 $\boldsymbol{\gamma}=\{\boldsymbol{\gamma}_q^g\}_{q=1,2,\cdots,r}^{g=1,2,\cdots,G}$。这样,式(2)所示的对偶优化问题可以改写为 $\max\limits_{\boldsymbol{\alpha}}\min\limits_{\boldsymbol{\gamma}}J(\boldsymbol{\alpha},\boldsymbol{\gamma})$。

$$J(\boldsymbol{\alpha},\boldsymbol{\gamma})=-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\boldsymbol{\alpha}_i^{\mathrm{T}}\boldsymbol{\alpha}_j\sum_{q=1}^{r}\gamma_q^{c(\boldsymbol{x}_i)}\gamma_q^{c(\boldsymbol{x}_j)}k_q(\boldsymbol{x}_i,\boldsymbol{x}_j)-\frac{1}{2C}\sum_{i=1}^{n}\boldsymbol{\alpha}_i^{\mathrm{T}}\boldsymbol{\alpha}_i+\sum_{i=1}^{n}\boldsymbol{\alpha}_i^{\mathrm{T}}\boldsymbol{Y}_i \tag{6}$$

式中:$c(\boldsymbol{x}_i)$ 和 $c(\boldsymbol{x}_j)$ 分别为样本 $\boldsymbol{x}_i$ 和 $\boldsymbol{x}_j$ 通过式(5)得到的聚类标签。

由式(6)可知,J 是一个关于 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\gamma}$ 的多目标函数。当 $\boldsymbol{\gamma}$ 固定时,关于 $\boldsymbol{\alpha}$ 最大化 J,意味着最小化全局分类误差;当 $\boldsymbol{\alpha}$ 固定时,关于 $\boldsymbol{\gamma}$ 最小化 J,意味着在最大化类内相似性的同时,最小化类间相似性。下面采用一种两步交替优化的策略,交替更新 ELM 模型参数 $\boldsymbol{\alpha}$ 和局部核权重 $\boldsymbol{\gamma}$。

2.2.1 在已知 $\boldsymbol{\gamma}$ 的条件下更新 $\boldsymbol{\alpha}$

当 $\boldsymbol{\gamma}$ 固定时,为了表示方便,本节采取一种矩阵向量化的操作。对于 Lagrange 乘子矩阵 $\boldsymbol{\alpha}\in\mathbf{R}^{n\times m}$,所谓的矩阵向量化记作 $\boldsymbol{\alpha}'=\psi(\boldsymbol{\alpha})$,即 $\boldsymbol{\alpha}'\in\mathbf{R}^{mn\times 1}$,且有 $\boldsymbol{\alpha}'=[\boldsymbol{\alpha}_1^{\mathrm{T}},\boldsymbol{\alpha}_2^{\mathrm{T}},\cdots,\boldsymbol{\alpha}_n^{\mathrm{T}}]^{\mathrm{T}}$。同理,理想输出矩阵 $\boldsymbol{Y}$ 经过矩阵向量化操作得到的新向量记作 $\boldsymbol{Y}'\in\mathbf{R}^{mn\times 1}$,且有 $\boldsymbol{Y}'=[\boldsymbol{Y}_1^{\mathrm{T}},\boldsymbol{Y}_2^{\mathrm{T}},\cdots,\boldsymbol{Y}_n^{\mathrm{T}}]^{\mathrm{T}}$。

组合核矩阵记作 $\boldsymbol{K}(\boldsymbol{\gamma})=\sum\limits_{q=1}^{r}(\boldsymbol{\gamma}_q\boldsymbol{\gamma}_q^{\mathrm{T}})\circ\boldsymbol{K}_q$。其中,$\boldsymbol{\gamma}_q=[\gamma_q^{c(\boldsymbol{x}_1)},\gamma_q^{c(\boldsymbol{x}_2)},\cdots,\gamma_q^{c(\boldsymbol{x}_n)}]^{\mathrm{T}}$,"$\circ$"表示矩阵之间对应元素分别相乘。进一步得到组合核矩阵 $\boldsymbol{K}(\boldsymbol{\gamma})$ 的一个增广矩阵,记作 $\boldsymbol{K}'\in\mathbf{R}^{mn\times mn}$,其具体形式如下:

$$\boldsymbol{K}'=\begin{bmatrix}\overbrace{\boldsymbol{K}(\boldsymbol{\gamma})_{11}}^{m} & \cdots & 0 & \cdots & \overbrace{\boldsymbol{K}(\boldsymbol{\gamma})_{1n}}^{m} & \cdots & 0\\ \vdots & \ddots & \vdots & \cdots & \vdots & \ddots & \vdots\\ 0 & \cdots & \boldsymbol{K}(\boldsymbol{\gamma})_{11} & \cdots & 0 & \cdots & \boldsymbol{K}(\boldsymbol{\gamma})_{1n}\\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots\\ \boldsymbol{K}(\boldsymbol{\gamma})_{n1} & \cdots & 0 & \cdots & \boldsymbol{K}(\boldsymbol{\gamma})_{nn} & \cdots & 0\\ \vdots & \ddots & \vdots & \cdots & \vdots & \ddots & \vdots\\ 0 & \cdots & \boldsymbol{K}(\boldsymbol{\gamma})_{n1} & \cdots & 0 & \cdots & \boldsymbol{K}(\boldsymbol{\gamma})_{nn}\end{bmatrix}\begin{matrix}\}m\\ \\ \}m\\ \\ \\ \}m\\ \\ \}m\end{matrix}$$

式中:$\boldsymbol{K}(\boldsymbol{\gamma})_{ij}$ 为 $\boldsymbol{K}(\boldsymbol{\gamma})$ 中第 i 行的第 j 个元素,且有 $\boldsymbol{K}(\boldsymbol{\gamma})_{ij}=\sum\limits_{q=1}^{r}\gamma_q^{c(\boldsymbol{x}_i)}\gamma_q^{c(\boldsymbol{x}_j)}k_q(\boldsymbol{x}_i,\boldsymbol{x}_j)$。

因此,式(6)所示的优化问题进一步被改写为 $\max\limits_{\boldsymbol{\alpha}'}J(\boldsymbol{\alpha}')$。

$$J(\boldsymbol{\alpha}')=-\frac{1}{2}\boldsymbol{\alpha}'^{\mathrm{T}}\boldsymbol{K}'\boldsymbol{\alpha}'-\frac{1}{2C}\boldsymbol{\alpha}'^{\mathrm{T}}\boldsymbol{\alpha}'+\boldsymbol{\alpha}'^{\mathrm{T}}\boldsymbol{Y}' \tag{7}$$

式(7)对 $\boldsymbol{\alpha}'$ 求偏导,并令结果等于 0,得到

$$\boldsymbol{\alpha}'=(\boldsymbol{K}'+\boldsymbol{I}/C)^{-1}\boldsymbol{Y}' \tag{8}$$

式中:$\boldsymbol{I}$ 为一个 $m\times n$ 阶单位矩阵。

最终,ELM 的模型参数可以通过一个矩阵向量化的反操作得到,记作 $\boldsymbol{\alpha}=\psi^{-1}(\boldsymbol{\alpha}')$。

2.2.2 在已知 $\boldsymbol{\alpha}$ 的条件下更新 $\boldsymbol{\gamma}$

结合聚类信息,式(6)所示的优化问题的目标函数等效于下面的形式:

$$J(\boldsymbol{\alpha},\boldsymbol{\gamma})=-\frac{1}{2}\sum_{s=1}^{G}\sum_{s'=1}^{G}\sum_{q=1}^{r}\gamma_q^s\gamma_q^{s'}\left(\sum_{i\mid c(\boldsymbol{x}_i)=s}\sum_{j\mid c(\boldsymbol{x}_j)=s'}\boldsymbol{\alpha}_i^{\mathrm{T}}\cdot\boldsymbol{\alpha}_jk_q(\boldsymbol{x}_i,\boldsymbol{x}_j)\right)-\frac{1}{2C}\sum_{i=1}^{n}\boldsymbol{\alpha}_i^{\mathrm{T}}\boldsymbol{\alpha}_i+\sum_{i=1}^{n}\boldsymbol{\alpha}_i^{\mathrm{T}}\boldsymbol{Y}_i$$

为了表示方便,不妨令

$$R_q^{ss'}(\boldsymbol{\alpha})=-\frac{1}{2}\sum_{i\mid c(\boldsymbol{x}_i)=s}\sum_{j\mid c(\boldsymbol{x}_j)=s'}\boldsymbol{\alpha}_i^{\mathrm{T}}\boldsymbol{\alpha}_jk_q(\boldsymbol{x}_i,\boldsymbol{x}_j) \tag{9}$$

$R_q^{ss}(\boldsymbol{\alpha})$ 体现了聚类 s 和聚类 s' 关于第 q 个基核的相关性。当 s 和 s' 相等时,$R_q^{ss'}(\boldsymbol{\alpha})$ 表示类内相似性;当 s 和 s' 不相等时,$R_q^{ss'}(\boldsymbol{\alpha})$ 表示类间相似性。相应地,优化问题式(6)被简写为 $\min\limits_{\boldsymbol{\gamma}}J(\boldsymbol{\gamma})$。

$$J(\boldsymbol{\gamma})=\sum_{s=1}^{G}\sum_{s'=1}^{G}\sum_{q=1}^{r}\gamma_q^s\gamma_q^{s'}R_q^{ss'}(\boldsymbol{\alpha})-\frac{1}{2C}\sum_{i=1}^{n}\boldsymbol{\alpha}_i^{\mathrm{T}}\boldsymbol{\alpha}_i+\sum_{i=1}^{n}\boldsymbol{\alpha}_i^{\mathrm{T}}\boldsymbol{Y}_i \tag{10}$$

式(10)中局部权重的优化是一个困难的二次非凸问题。受文献[16]的启发,本节从训练数据在输入空间和特征空间的统计特性出发,提出 2 种不同的选通函数。

1) 函数 M_1:面向输入空间的选通函数

在输入空间中,第 g 个聚类在第 q 个基核上的选通函数定义为

http://bhxb.buaa.edu.cn

$$\eta_q^g(\boldsymbol{\mu}_q,\sigma_q)=\frac{\exp(-\|\boldsymbol{e}_g-\boldsymbol{\mu}_q\|_2^2/\sigma_q^2)}{\sum_{h=1}^{r}\exp(-\|\boldsymbol{e}_g-\boldsymbol{\mu}_h\|_2^2/\sigma_h^2)} \tag{11}$$

式中:$\{\boldsymbol{\mu}_h,\sigma_h\}_{h=1}^{r}$ 为选通模型参数;$\boldsymbol{e}_g$ 为通过 2.1 节计算得到的第 g 个聚类的聚类中心。令 $\gamma_q^g\approx\eta_q^g(\boldsymbol{\mu}_q,\sigma_q)$,在 γ_q^g 中分别对模型参数 $\boldsymbol{\mu}_p,\sigma_p$ 求偏导,可以得到

$$\frac{\partial\gamma_q^g}{\partial\boldsymbol{\mu}_p}=\begin{cases}-\dfrac{2(\boldsymbol{e}_g-\boldsymbol{\mu}_p)}{\sigma_p^2}\gamma_p^g\gamma_q^g & q\neq p\\ \dfrac{2(\boldsymbol{e}_g-\boldsymbol{\mu}_p)}{\sigma_p^2}\gamma_p^g-\dfrac{2(\boldsymbol{e}_g-\boldsymbol{\mu}_p)}{\sigma_p^2}(\gamma_p^g)^2 & q=p\end{cases}$$

$$\frac{\partial\gamma_q^g}{\partial\sigma_p}=\begin{cases}-\dfrac{2\|\boldsymbol{e}_g-\boldsymbol{\mu}_p\|_2^2}{\sigma_p^3}\gamma_p^g\gamma_q^g & q\neq p\\ \dfrac{2\|\boldsymbol{e}_g-\boldsymbol{\mu}_p\|_2^2}{\sigma_p^3}\gamma_p^g-\dfrac{2\|\boldsymbol{e}_g-\boldsymbol{\mu}_p\|_2^2}{\sigma_p^3}(\gamma_p^g)^2 & q=p\end{cases}$$

基于上述结果,式(10)中的 $J(\boldsymbol{\gamma})$ 对 $\boldsymbol{\mu}_p,\sigma_p$ 分别求偏导,进一步得到

$$\begin{aligned}\frac{\partial J(\boldsymbol{\gamma})}{\partial\boldsymbol{\mu}_p}=&\sum_{s=1}^{G}\sum_{s'=1}^{G}\sum_{q=1,q\neq p}^{r}\left[\left(\frac{\partial\gamma_q^s}{\partial\boldsymbol{\mu}_p}\right)\gamma_q^{s'}+\gamma_q^s\left(\frac{\partial\gamma_q^{s'}}{\partial\boldsymbol{\mu}_p}\right)\right]R_q^{ss'}(\boldsymbol{\alpha})+\sum_{s=1}^{G}\sum_{s'=1}^{G}\left[\left(\frac{\partial\gamma_p^s}{\partial\boldsymbol{\mu}_p}\right)\gamma_p^{s'}+\gamma_p^s\left(\frac{\partial\gamma_p^{s'}}{\partial\boldsymbol{\mu}_p}\right)\right]R_p^{ss'}(\boldsymbol{\alpha})=2\sum_{s=1}^{G}\sum_{s'=1}^{G}\sum_{q=1}^{r}\gamma_q^s\gamma_q^{s'}R_q^{ss'}(\boldsymbol{\alpha})\cdot\left[\frac{(\boldsymbol{e}_s-\boldsymbol{\mu}_p)(\delta_p^q-\gamma_p^s)}{\sigma_p^2}\right]+\\&2\sum_{s=1}^{G}\sum_{s'=1}^{G}\sum_{q=1}^{r}\gamma_q^s\gamma_q^{s'}R_q^{ss'}(\boldsymbol{\alpha})\cdot\left[\frac{(\boldsymbol{e}_{s'}-\boldsymbol{\mu}_p)(\delta_p^q-\gamma_p^{s'})}{\sigma_p^2}\right]\end{aligned} \tag{12}$$

$$\begin{aligned}\frac{\partial J(\boldsymbol{\gamma})}{\partial\sigma_p}=&\sum_{s=1}^{G}\sum_{s'=1}^{G}\sum_{q=1,q\neq p}^{r}\left[\left(\frac{\partial\gamma_q^s}{\partial\sigma_p}\right)\gamma_q^{s'}+\gamma_q^s\left(\frac{\partial\gamma_q^{s'}}{\partial\sigma_p}\right)\right]R_q^{ss'}(\boldsymbol{\alpha})+\sum_{s=1}^{G}\sum_{s'=1}^{G}\left[\left(\frac{\partial\gamma_p^s}{\partial\sigma_p}\right)\gamma_p^{s'}+\gamma_p^s\left(\frac{\partial\gamma_p^{s'}}{\partial\sigma_p}\right)\right]R_p^{ss'}(\boldsymbol{\alpha})=2\sum_{s=1}^{G}\sum_{s'=1}^{G}\sum_{q=1}^{r}\gamma_q^s\gamma_q^{s'}R_q^{ss'}(\boldsymbol{\alpha})\cdot\left[\frac{\|\boldsymbol{e}_s-\boldsymbol{\mu}_p\|_2^2(\delta_p^q-\gamma_p^s)}{\sigma_p^3}\right]+\\&2\sum_{s=1}^{G}\sum_{s'=1}^{G}\sum_{q=1}^{r}\gamma_q^s\gamma_q^{s'}R_q^{ss'}(\boldsymbol{\alpha})\cdot\left[\frac{\|\boldsymbol{e}_{s'}-\boldsymbol{\mu}_p\|_2^2(\delta_p^q-\gamma_p^{s'})}{\sigma_p^3}\right]\end{aligned} \tag{13}$$

式中:δ_p^q 为一个单位冲激函数,当 $q=p$ 时,$\delta_p^q=1$,否则 $\delta_p^q=0$。

得到目标函数 $J(\boldsymbol{\gamma})$ 关于模型参数的梯度之后,采用梯度下降法依次更新各模型参数,即

$$\begin{cases}\boldsymbol{\mu}_p(t+1)=\boldsymbol{\mu}_p(t)-\tau_{11}(t)\dfrac{\partial J(\boldsymbol{\gamma})}{\partial\boldsymbol{\mu}_p}\\ \sigma_p(t+1)=\sigma_p(t)-\tau_{12}(t)\dfrac{\partial J(\boldsymbol{\gamma})}{\partial\sigma_p}\end{cases} \tag{14}$$

式中:$\tau_{11}(t)$ 和 $\tau_{12}(t)$ 为迭代步长,在迭代过程中采用线性搜索算法得到。

2)函数 M_2:向特征空间的选通函数

在特征空间中,第 g 个聚类在第 q 个基核上的选通函数定义为

$$\eta_q^g(a_q^g,b_q^g)=\frac{\exp(a_q^gK_q^g+b_q^g)}{\sum_{h=1}^{r}\exp(a_h^gK_h^g+b_h^g)} \tag{15}$$

式中:$\{a_h^g,b_h^g\}_{h=1,2,\cdots,r}^{g=1,2,\cdots,G}$ 为选通函数的模型参数;K_q^g 为特征空间中第 g 个聚类与第 q 个核函数相关的统计特性。用 n_g 为第 g 个聚类包含的样本数量,则 K_q^g 定义为

$$K_q^g=\frac{\sum_{\{i\mid c(\boldsymbol{x}_i)=g\}}\sum_{\{j\mid c(\boldsymbol{x}_j)=g\}}k_q(\boldsymbol{x}_i,\boldsymbol{x}_j)}{n_g^2} \tag{16}$$

令 $\gamma_q^g\approx\eta_q^g(a_q^g,b_q^g)$,在 γ_q^g 中分别对模型参数 a_p^g,b_p^g 求偏导,可以得到

$$\frac{\partial\gamma_q^g}{\partial a_p^g}=\begin{cases}-K_p^g\gamma_p^g\gamma_q^g & q\neq p\\ K_p^g\gamma_p^g-K_p^g(\gamma_p^g)^2 & q=p\end{cases}$$

$$\frac{\partial\gamma_q^g}{\partial b_p^g}=\begin{cases}-\gamma_p^g\gamma_q^g & q\neq p\\ \gamma_p^g-(\gamma_p^g)^2 & q=p\end{cases}$$

基于以上结果,式(10)中的 $J(\boldsymbol{\gamma})$ 对 a_p^g,b_p^g 分别求偏导,进一步得到

$$\begin{aligned}\frac{\partial J(\boldsymbol{\gamma})}{\partial a_p^g}=&\frac{\partial}{\partial a_p^g}\sum_{s=1}^{G}\sum_{s'=1}^{G}\sum_{q=1}^{r}\gamma_q^s\gamma_q^{s'}R_q^{ss'}(\boldsymbol{\alpha})=\sum_{s=1,s\neq g}^{G}\sum_{q=1,q\neq p}^{r}\gamma_q^s\left(\frac{\partial\gamma_q^g}{\partial a_p^g}\right)R_q^{sg}(\boldsymbol{\alpha})+\sum_{s=1,s\neq g}^{G}\gamma_p^s\left(\frac{\partial\gamma_p^g}{\partial a_p^g}\right)R_p^{sg}(\boldsymbol{\alpha})+\\&\sum_{q=1,q\neq p}^{r}\sum_{s'=1,s'\neq g}^{G}\left(\frac{\partial\gamma_q^g}{\partial a_p^g}\right)\gamma_q^{s'}R_q^{gs'}(\boldsymbol{\alpha})+\sum_{q=1,q\neq p}^{r}\left(\frac{\partial\gamma_q^g}{\partial a_p^g}\right)2\gamma_q^gR_q^{gg}(\boldsymbol{\alpha})+\\&\sum_{s'=1,s'\neq g}^{G}\left(\frac{\partial\gamma_p^g}{\partial a_p^g}\right)\gamma_p^{s'}R_p^{gs'}(\boldsymbol{\alpha})+\left(\frac{\partial\gamma_p^g}{\partial a_p^g}\right)2\gamma_p^gR_p^{gg}(\boldsymbol{\alpha})\end{aligned}$$

$$\begin{aligned}\frac{\partial J(\boldsymbol{\gamma})}{\partial b_p^g}=&\frac{\partial}{\partial b_p^g}\sum_{s=1}^{G}\sum_{s'=1}^{G}\sum_{q=1}^{r}\gamma_q^s\gamma_q^{s'}R_q^{ss'}(\boldsymbol{\alpha})=\sum_{s=1,s\neq g}^{G}\sum_{q=1,q\neq p}^{r}\gamma_q^s\left(\frac{\partial\gamma_q^g}{\partial b_p^g}\right)R_q^{sg}(\boldsymbol{\alpha})+\\&\sum_{s=1,s\neq g}^{G}\gamma_p^s\left(\frac{\partial\gamma_p^g}{\partial b_p^g}\right)R_p^{sg}(\boldsymbol{\alpha})+\end{aligned}$$

http://bhxb.buaa.edu.cn

$$\sum_{q=1,q\neq p}^{r}\sum_{s'=1,s'\neq g}^{G}\left(\frac{\partial\gamma_q^g}{\partial b_p^g}\right)\gamma_q^{s'}R_q^{gs'}(\boldsymbol{\alpha})+$$

$$\sum_{q=1,q\neq p}^{r}\left(\frac{\partial\gamma_q^g}{\partial b_p^g}\right)2\gamma_q^gR_q^{gg}(\boldsymbol{\alpha})+$$

$$\sum_{s'=1,s'\neq g}^{G}\left(\frac{\partial\gamma_p^g}{\partial b_p^g}\right)\gamma_p^{s'}R_p^{gs'}(\boldsymbol{\alpha})+\left(\frac{\partial\gamma_p^g}{\partial b_p^g}\right)2\gamma_p^gR_p^{gg}(\boldsymbol{\alpha})$$

由式(9)可知,$R_q^{ss'}(\boldsymbol{\alpha})$关于 s 和 s'是对称的。所以,基于上述结果可得

$$\frac{\partial J(\boldsymbol{\gamma})}{\partial a_p^g}=2\sum_{q=1}^{r}\left(\sum_{s=1}^{G}K_p\gamma_p^g(\delta_p^q-\gamma_q^g)\gamma_q^sR_q^{sg}(\boldsymbol{\alpha})\right)\tag{17}$$

同理得到

$$\frac{\partial J(\boldsymbol{\gamma})}{\partial b_p^g}=2\sum_{q=1}^{r}\left(\sum_{s=1}^{G}\gamma_p^g(\delta_p^q-\gamma_q^g)\gamma_q^sR_q^{sg}(\boldsymbol{\alpha})\right)\tag{18}$$

得到目标函数 $J(\boldsymbol{\gamma})$关于模型参数的梯度之后,同样采用梯度下降法更新各模型参数,即

$$\begin{cases}a_p^g(t+1)=a_p^g(t)-\tau_{21}(t)\dfrac{\partial J(\boldsymbol{\gamma})}{\partial a_p^g}\\ b_p^g(t+1)=b_p^g(t)-\tau_{22}(t)\dfrac{\partial J(\boldsymbol{\gamma})}{\partial b_p^g}\end{cases}\tag{19}$$

式中:$\tau_{21}(t)$和 $\tau_{22}(t)$为迭代步长,在迭代过程中采用线性搜索算法得到。

2.3 诊断决策

在诊断过程中,给定一个测试数据实例 $\boldsymbol{x}'_j$,本文模型首先根据 AP 聚类找出它所属的群组 $\boldsymbol{X}_s(s=1,2,\cdots,G)$;然后将群组 $\boldsymbol{X}_s$ 的局部核权重 $\{\gamma_q^s\}_{q=1}^r$分配给样本 $\boldsymbol{x}'_j$。假设通过 2 步交替优化过程,得到的最优的诊断模型参数为 $\boldsymbol{\alpha}^*$ 和 $\boldsymbol{\gamma}^*$,则根据式(3),本文模型的决策函数定义为

$$f(\boldsymbol{x}'_j)=\sum_{g=1}^{G}\sum_{q=1}^{r}(\boldsymbol{\gamma}^*)_q^s(\boldsymbol{\gamma}^*)_q^g\cdot\left(\sum_{|i|c(\boldsymbol{x}_i)=g\}}(\boldsymbol{\alpha}_i^*)^{\mathrm{T}}k_q(\boldsymbol{x}_i,\boldsymbol{x}'_j)\right)\tag{20}$$

决策过程如图 2 所示。

对于 LCMKELM 诊断模型而言,其输出节点个数 m 等于故障模式的个数 c。如果第 i 个样本对应的故障模式是 l,那么分类器的理想输出应该为 $\boldsymbol{Y}_i=[0,\cdots,0,\underset{l}{1},0,\cdots,0]^{\mathrm{T}}$。也就是说在输出向量 $\boldsymbol{Y}_i$ 中仅仅只有第 l 个元素为 1。不妨设 $f(\cdot)=\{f^{(1)}(\cdot),f^{(2)}(\cdot),\cdots,f^{(c)}(\cdot)\}$,其中 $f^{(l)}(\cdot)$为诊断模型第 l 个输出节点的输出值,则测试样本 $\boldsymbol{x}'_j$ 对应的故障模式被判别为

$$y'_j=\arg\max_{l\in\{1,2,\cdots,c\}}f^{(l)}(\boldsymbol{x}'_j)\tag{21}$$

本文故障诊断模型流程图如图 3 所示。

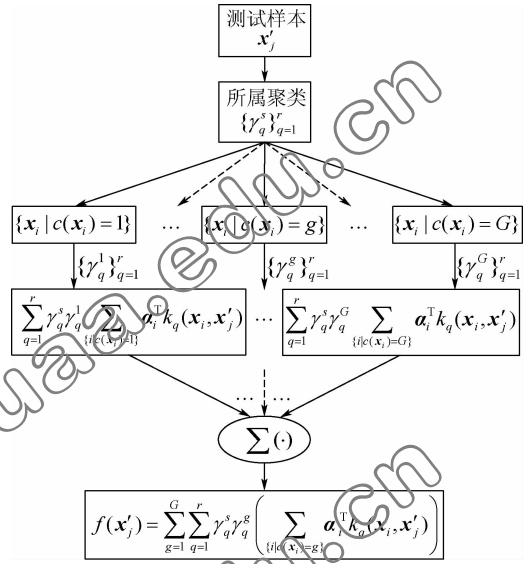


图 2 本文模型的决策过程

Fig. 2 Decision making process of proposed model

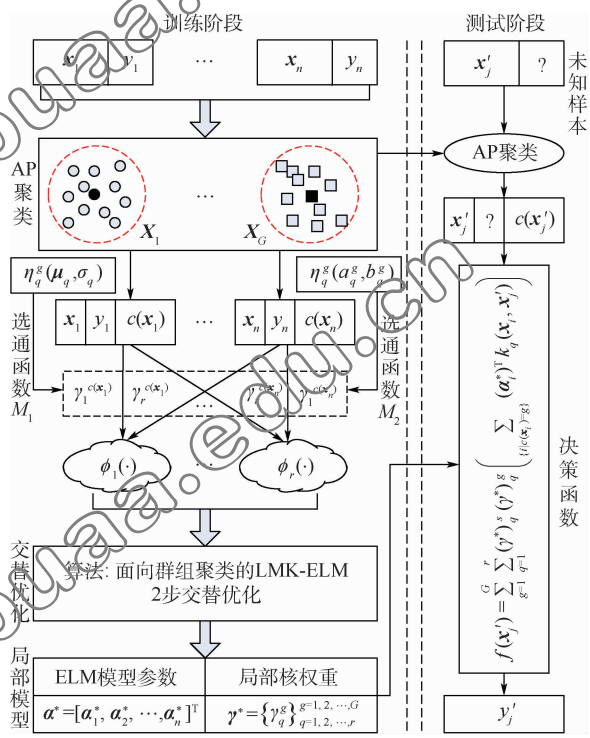


图 3 本文故障诊断模型流程图

Fig. 3 Flowchart of proposed fault diagnosis model

3 算法流程

为表示方便,将本文诊断模型记作 LCMKELM。当采用不同选通函数时,分别记作 M_1-LCMKELM 和 M_2-LCMKELM。根据图 3 将诊断模型的详细流程总结如下:

步骤 1 输入训练数据集 $\boldsymbol{D}_{\mathrm{Tr}}=\{(\boldsymbol{x}_1,y_1),(\boldsymbol{x}_2,y_2),\cdots,(\boldsymbol{x}_i,y_i),\cdots,(\boldsymbol{x}_n,y_n)\}$;设置基核

http://bhxb.buaa.edu.cn

$\{k_q(\cdot,\cdot)\}_{q=1}^{r}$;AP 算法衰减系数 λ,最大迭代次数以及 ELM 正则化因子 C。

步骤 2　通过式(4)构建训练样本相似度矩阵 $\boldsymbol{S}$,并基于 AP 算法对训练样本进行聚类划分 $\{\boldsymbol{X}_g\}_{g=1}^{G}$;若采用选通函数 M_1,初始化模型参数 $\{\boldsymbol{\mu}_q,\boldsymbol{\sigma}_q\}_{q=1}^{r}$,基于聚类划分提取聚类中心 $\{\boldsymbol{e}_g\}_{g=1}^{G}$;若采用选通函数 M_2,初始化模型参数 $\{a_q^g,b_q^g\}_{q=1,2,\cdots,r}^{g=1,2,\cdots,G}$;基于式(16)计算统计量 $\{K_q^g\}_{q=1,2,\cdots,r}^{g=1,2,\cdots,G}$。

步骤 3　若采用选通函数 M_1,令 $\gamma_q^g\approx\eta_q^g(\boldsymbol{\mu}_q,\boldsymbol{\sigma}_q)$,并计算 $\{\gamma_q^g\}_{q=1,2,\cdots,r}^{g=1,2,\cdots,G}$;若采用选通函数 M_2,令 $\gamma_q^g\approx\eta_q^g(a_q^g,b_q^g)$,并计算 $\{\gamma_q^g\}_{q=1,2,\cdots,r}^{g=1,2,\cdots,G}$。

步骤 4　基于 $\{\gamma_q^g\}_{q=1,2,\cdots,r}^{g=1,2,\cdots,G}$ 计算组合核矩阵 $\boldsymbol{K}(\boldsymbol{\gamma})$,并计算 $\boldsymbol{K}(\boldsymbol{\gamma})$ 的增广矩阵 $\boldsymbol{K}'$;同时,计算矩阵向量化后的理想输出向量 $\boldsymbol{Y}'$。

步骤 5　通过式(8)计算 $\boldsymbol{\alpha}'$,并进一步得到 ELM 模型参数 $\boldsymbol{\alpha}$;同时,通过式(9)计算 $R_q^{ss'}(\boldsymbol{\alpha})$。

步骤 6　令 $t=1$,通过式(10)计算 $J_t(\boldsymbol{\gamma})$。

步骤 7　若采用选通函数 M_1,通过式(12)和式(13)分别计算 $\dfrac{\partial J(\boldsymbol{\gamma})}{\partial\boldsymbol{\mu}_p}$ 和 $\dfrac{\partial J(\boldsymbol{\gamma})}{\partial\sigma_p}$,$p=1,2,\cdots,r$;通过式(14)更新 $\boldsymbol{\mu}_p,\sigma_p$;若选择选通函数 M_2,通过式(17)和式(18)分别计算 $\dfrac{\partial J(\boldsymbol{\gamma})}{\partial a_p^g}$ 和 $\dfrac{\partial J(\boldsymbol{\gamma})}{\partial b_p^g}$,$p=1,2,\cdots,r$;$g=1,2,\cdots,G$;通过式(19)更新 a_p^g,b_p^g。

步骤 8　执行步骤 3~步骤 5;通过式(10)计算 $J_{t+1}(\boldsymbol{\gamma})$;如果 $|J_{t+1}(\boldsymbol{\gamma})-J_t(\boldsymbol{\gamma})|>10^{-3}$,$t=t+1$,返回步骤 7;否则,执行步骤 9。

步骤 9　输入测试样本实例 $\boldsymbol{x}_j'$,$j=1,2,\cdots,n'$;通过式(4)计算 $\boldsymbol{x}_j'$ 与每个聚类中心的相似度,进而确定 $\boldsymbol{x}_j'$ 所属的群组 $s\in\{1,2,\cdots,G\}$;根据 s 为 $\boldsymbol{x}_j'$ 分配局部核权重 $\{\gamma_q^s\}_{q=1}^{r}$。

步骤 10　通过式(20)计算决策模型输出值 $f(\boldsymbol{x}_j')$;通过式(21)得到测试数据实例 $\boldsymbol{x}_j'$ 对应的故障模式 y_j'。

4　实验分析

本节通过 2 个实验来验证本文模型的性能。SimpleMKL[13]、GMKL-SVM[12]和 LMKL[16]被采用作为比较算法,对于多分类问题,其均采用“一对一”的原则。在实验中所有的数据首先进行 Z-score标准化处理,而所有的核矩阵则被正规化到具有单位迹,即 $\mathrm{tr}(\boldsymbol{K}_q)=1$。正则化参数 C 采用 5 倍交叉验证从集合 $\{10^{-2},0.1,1,10,10^2,10^3,10^4\}$ 中选择。AP 算法中令 $\lambda=0.8$,最大迭代次数设置为 1 000。算法复杂性通过训练时间(不包括交叉验证时间)和测试时间共同评价。所有实验结果通过 MATLAB 2015a 得到,实验电脑配置为:2.27 GHz Intel Core i3 CPU,2 GB RAM。

4.1　算法有效性验证

本节采用文献[16]中使用的人工数据集 Gauss 4 来验证本文模型的有效性。该数据集由服从 4 种不同高斯分布的 1 200 个样本组成,包括 2 个类别,每个类别包括 2 种分布。每种分布的先验概率、均值向量和协方差矩阵分别为

$$\rho_{11}=\rho_{12}=\rho_{21}=\rho_{22}=0.25 \tag{22}$$

$$\begin{cases}\boldsymbol{o}_{11}=\begin{bmatrix}-3.0\\1.0\end{bmatrix}\\ \boldsymbol{o}_{12}=\begin{bmatrix}1.0\\1.0\end{bmatrix}\\ \boldsymbol{o}_{21}=\begin{bmatrix}-1.0\\-2.2\end{bmatrix}\\ \boldsymbol{o}_{22}=\begin{bmatrix}3.0\\-2.2\end{bmatrix}\end{cases} \tag{23}$$

$$\begin{cases}\boldsymbol{\varepsilon}_{11}=\begin{bmatrix}0.8&0\\0&2.0\end{bmatrix}\\ \boldsymbol{\varepsilon}_{12}=\begin{bmatrix}0.8&0\\0&2.0\end{bmatrix}\\ \boldsymbol{\varepsilon}_{21}=\begin{bmatrix}0.8&0\\0&4.0\end{bmatrix}\\ \boldsymbol{\varepsilon}_{22}=\begin{bmatrix}0.8&0\\0&4.0\end{bmatrix}\end{cases} \tag{24}$$

实验中,随机选择 2/3 数据作为训练样本,剩余 1/3 作为测试样本。首先采用 AP 算法对训练样本进行聚类,800 组训练样本被划分为 26 个群组,聚类结果如图 4 所示。

SimpleMKL 被用来与本文模型进行比较。多核设置为线性核,多项式核(参数为 2),高斯核(参数为 1)。评价指标采用文献[22]中定义的分类精度、F1 值和 G-mean。实验共进行 5 次,将所得结果的均值和标准差记录于表 1 中。

由表 1 的结果看到,基于选通函数 M_1、M_2 的 LCMKELM 均可以实现比 SimpleMKL 更优的分类性能。其克服了局部算法将基核权重拟合到每一个确定样本上时可能出现的过学习问题。对于分类精度,相比于 SimpleMKL,M_1-LCMKELM 和 M_2-LCMKELM 分别提升了 1.65% 和 1.85%。

为了更加直观地比较本文模型的性能,将其与 SimpleMKL 的接受者操作特性曲线(Receiver Operating characteristic Curve, ROC)绘制在图 5

http://bhxb.buaa.edu.cn

中。由图 5 可见,2 种算法的曲线下方区域(Area Under the Curve, AUC)明显大于 SimpleMKL 的 AUC 值,这从另一个角度进一步证明了本文模型的有效性。

为验证训练样本数量对模型精度的影响。设置训练样本数量分别等于 16,32,48,64,144,400,600,800。每次随机选择 100 组测试样本,将其注入到训练得到的不同模型中。实验共进行 5 次,将所得结果的均值与标准差记录于表 2 中。

由表 2 看到,训练样本的增加确实提升了诊断模型的分类性能。但是 2 种算法通过挖掘样本中蕴含的局部特征信息,在小样本条件下可以得到与较大样本条件下相近的分类精度。该结果证明了将本文模型应用于小样本条件下故障诊断任务的可行性。

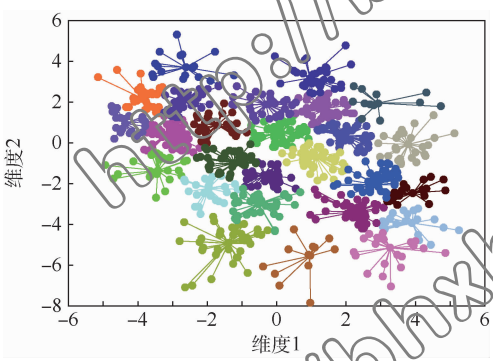


图 4 Gauss 4 训练样本 AP 算法聚类结果

Fig. 4 Clustering results of Gauss 4 training samples based on AP algorithm

表 1 Gauss 4 数据集分类结果比较

Table 1 Comparison of classification results on Gauss 4 dataset

评价指标	M_1-LCMKELM	M_2-LCMKELM	SimpleMKL
分类精度/%	90.500 0 ± 0.684 7	90.850 0 ± 0.602 1	89.350 0 ± 1.206 8
F1 值	0.904 9 ± 0.006 8	0.908 5 ± 0.006 1	0.895 8 ± 0.012 4
G-mean	0.904 6 ± 0.006 9	0.908 2 ± 0.006 2	0.893 2 ± 0.011 9

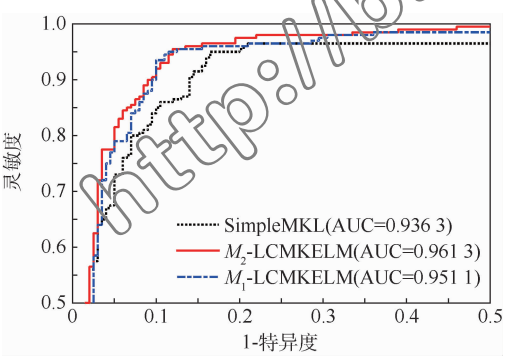


图 5 Gauss 4 数据集 ROC 曲线比较

Fig. 5 Comparison of ROC curves on Gauss 4 dataset

表 2 不同训练样本数量下分类精度比较

Table 2 Comparison of classification accuracy with different sizes of training sample

训练样本数量	分类精度/%	
	M_1-LCMKELM	M_2-LCMKELM
16	81.8 ± 1.923 5	82.6 ± 1.673 3
32	85.2 ± 1.483 2	86.2 ± 1.483 2
48	86.4 ± 2.073 6	86.6 ± 1.516 5
64	86.4 ± 1.516 5	86.6 ± 1.673 3
144	87.2 ± 1.095 4	87.4 ± 1.140 2
400	87.0 ± 1.870 8	87.2 ± 1.303 8
600	87.8 ± 1.095 4	88.2 ± 0.836 7
800	88.0 ± 0.707 1	88.6 ± 0.894 4

4.2 旋转变压器激励发生电路诊断实例

旋转变压器激励发生电路如图 6 所示。它用于产生旋转变压器磁绕组所需的正弦信号,为旋转变压器运行提供激励。以满足无刷直流电机、航空永磁同步电机等系统检测需求。该电路主要由电源模块、频率控制模块、正弦信号产生模块和幅值调理及驱动能力调节模块组成。

在基于自动测试系统(Automatic Test System, ATS)的测试流程中共包含 9 个测试项目:信号频率(Q_1),信号幅值(Q_2),正弦模块输入电压值(Q_3),信号频率稳定度(Q_4),+15 V 电压值(Q_5),-15 V 电压值(Q_6),电路板工作温度(Q_7),+5 V 电压值(Q_8),+10 V 电压值(Q_9)。用 F_0,F_1,F_2,F_3 及 F_4 分别表示正常模式、频率控制模块故障、幅值调理及驱动能力调节模块故障、电源模块故障和正弦信号产生模块故障。各模式下 ATS 共采集原始数据量分别为 22,18,16,14 及 22。

将采集到的数据随机选择 1/2 作为训练样本,其余作为测试样本。多核设置为:线性核,多项式核(参数 2),高斯核(参数 2,10,20,30,40,50)。首先采用 AP 算法对训练样本进行聚类,46 组训练样本被划分为 8 个群组,如表 3 所示。

用本文模型与 SimpleMKL、GMKL-SVM、LMKL-softmax 和 LMKL-sigmoid 分别基于上述数据进行诊断测试,所得结果如图 7 和表 4 所示。

由图 7 和表 4 可以得到:①本文算法在漏警率与虚警率上表现出色,实现了 0 漏警,0 虚警。②2 种LMKL 算法尽管具有 100% 的训练诊断精度,但其测试诊断精度不高,尤其是当采用 softmax 选通函数时。究其原因在于将局部权重拟合到确定样本上时导致算法泛化性能下降。而本文算法通过聚类将局部权重拟合到一组具有相似局部特征的群组上,在挖掘局部特征的同时避

http://bhxb.buaa.edu.cn

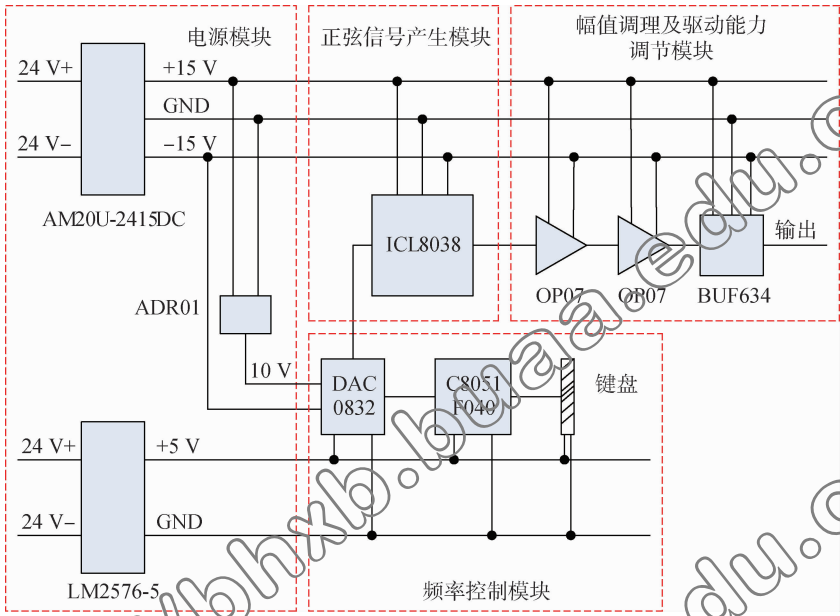


图 6　旋转变压器激励发生电路原理图

Fig. 6　Schematic diagram of rotary transformer excitation generating circuit

表 3　诊断数据聚类结果

Table 3　Clustering results of diagnosis data

聚　类	聚类中心	类内元素
1	5	{2,4,5,6,9,11,15,18,34,35}
2	12	{8,10,12,13,20,25}
3	17	{1,3,7,17}
4	23	{16,23,33}
5	26	{14,19,26,29,32}
6	30	{24,30,31,41}
7	40	{21,22,28,37,38,40,42,45}
8	43	{27,36,39,43,44,46}

免了算法过学习。相比于其他 4 种算法（SimpleMKL、GMKL-SVM、LMKL-softmax 和 LMKL-sigmoid），M_1-LCMKELM 将测试诊断精度分别提升了 4.35%、2.17%、6.52% 和 2.17%；M_2-LCMKELM 将测试诊断精度分别提升了 6.52%、4.35%、8.70% 和 4.35%。③比较发现 M_2-LCMKELM 在诊断精度上要优于 M_1-LCMKELM，原因在于当使用选通函数 M_1 时产生局部权重的稀疏解，当使用选通函数 M_2 时产生局部权重的非稀疏解，如图8所示。稀疏解虽然易于解释，但

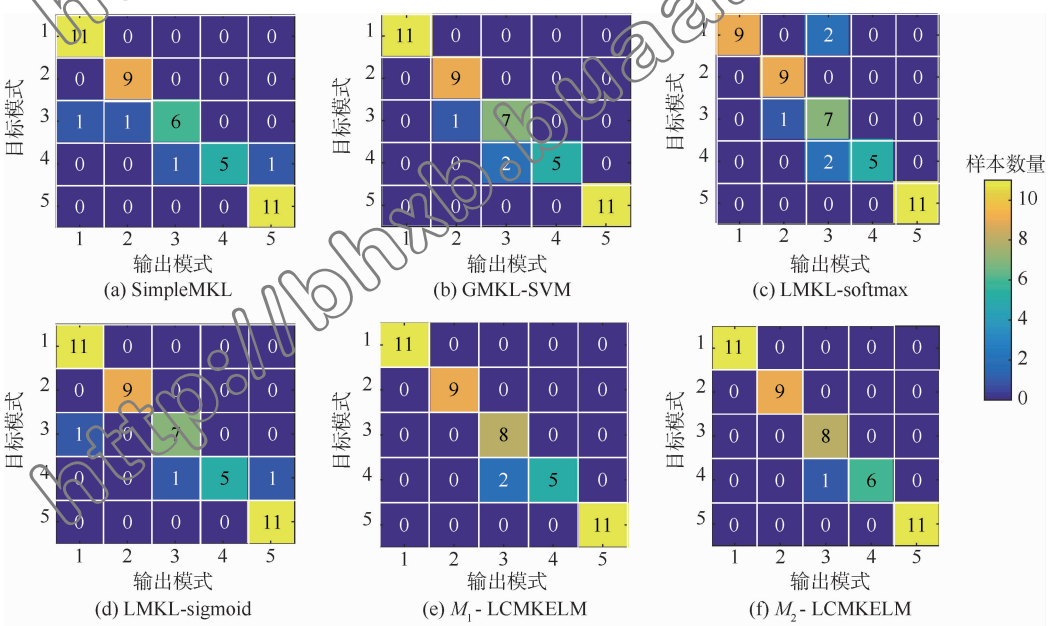


图 7　不同算法的混淆矩阵

Fig. 7　Confusion matrices of different algorithms

http://bhxb.buaa.edu.cn

可能会丢失有用的信息。

不妨用TP表示某种故障模式被正确识别的数量,FN表示该故障模式被错误识别的数量,FP表示误诊断为该故障模式的样本数量。则将召回率定义为TP/(TP+FN),而准确率定义为TP/(TP+FP)。因此,根据图7得到4种故障模式的召回率、准确率直方图,如图9所示。

根据文献[23]中关于多分类问题F1值及G-mean的定义,结合图9得到表5所示的诊断结果统计特性。表5从诊断精度之外的另一个角度展示了各种诊断算法的性能。可以看到本文算法依旧展示出更好的诊断性能。

此外,为了比较各种算法的时间花费,重复运行实验5次,将各种算法训练时间与测试时间的均值与标准差记录在表6中。

由表6可知,①在训练时间上,采用LMKC明显要比一般的MKL花费更多的时间,而所提的2种局部MKL算法与2种流行的局部MKL算法相比较,拥有相似、甚至更短的训练时间。②在测试时间上,本文算法花费时间最少,尤其是M_1-

表4 不同算法诊断结果比较

Table 4 Comparison of diagnosis results based on different algorithms

算法	漏警率/%	虚警率/%	训练诊断精度/%	测试诊断精度/%
SimpleMKL	2.857 1	0	100	91.304 3
GMKL-SVM	0	0	100	93.478 3
LMKL-softmax	0	5.405 4	100	89.130 4
LMKL-sigmoid	2.857 1	0	100	93.478 3
M_1-LCMKELM	0	0	100	95.652 2
M_2-LCMKELM	0	0	100	97.826 1

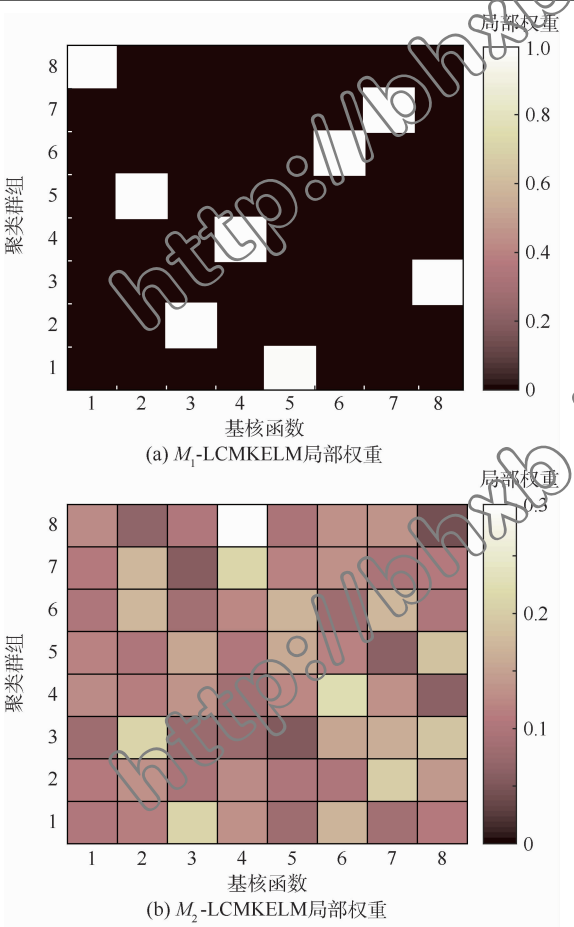
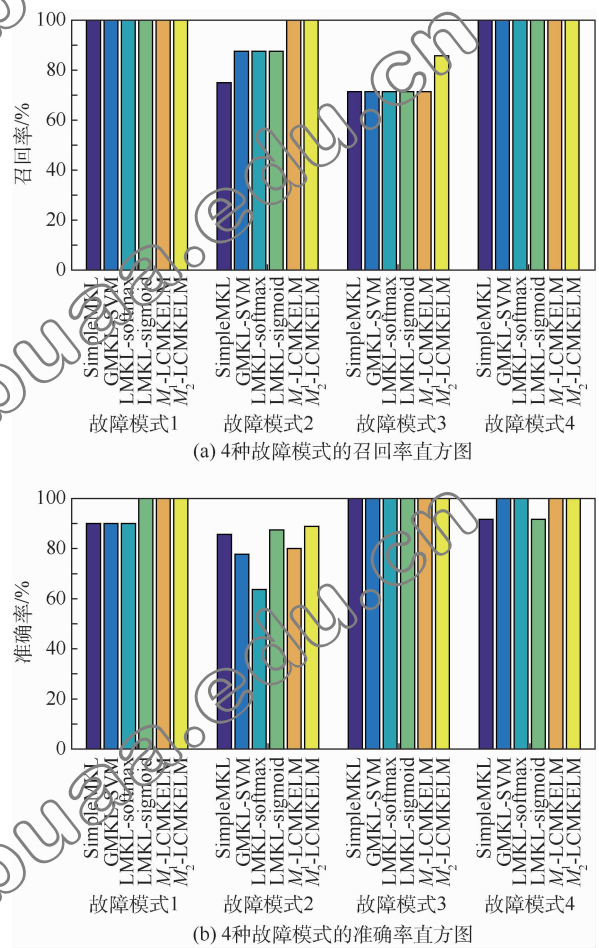


图8 本文算法局部权重分布

Fig. 8 Localized weight distribution of proposed algorithms



(a) 4种故障模式的召回率直方图

(b) 4种故障模式的准确率直方图

图9 不同算法的召回率、准确率比较

Fig. 9 Comparison of recall and precision based on different algorithms

表5 不同算法F1值和G-mean比较

Table 5 Comparison of F1 score and G-mean based on different algorithms

算法	F1值	G-mean
SimpleMKL	0.898 7	0.882 6
GMKL-SVM	0.920 8	0.910 3
LMKL-softamx	0.885 5	0.874 5
LMKL-sigmoid	0.924 3	0.910 3
M_1-LCMKELM	0.944 4	0.934 9
M_2-LCMKELM	0.972 9	0.969 6

http://bhxb.buaa.edu.cn

表 6　不同算法诊断时间花费比较

Table 6　Comparison of diagnosis time cost based on different algorithms

算　法	训练时间/s	测试时间/s
SimpleMKL	0.991 7 ± 0.082 9	0.201 8 ± 0.045 1
GMKL-SVM	0.832 5 ± 0.036 0	0.167 9 ± 0.008 2
LMKL-softamx	1.493 4 ± 0.062 5	0.173 5 ± 0.011 1
LMKL-sigmoid	2.715 9 ± 0.129 2	0.176 9 ± 0.015 7
M_1-LCMKELM	2.497 9 ± 0.213 7	0.129 9 ± 0.005 9
M_2-LCMKELM	1.789 7 ± 0.201 4	0.152 1 ± 0.016 7

LCMKELM。这是因为 ELM 相比于 SVM 本身就是一种计算更加高效的分类器,同时当采用输入空间选通函数 M_1 时,能够产生局部权重的稀疏解。③本文提出的是一种线下故障诊断算法,通过牺牲一定的时间花费获得诊断精度上的提升是可以接受的;况且,本文主要面向于小样本条件下的诊断问题,因此时间花费也是可控的。

5　结　论

针对小样本条件下,基于 ATS 测试数据的航空电子部件功能故障诊断问题,本文提出一种 LCMKELM 诊断模型。通过实验验证了所提模型的有效性和适用性,结果表明:

1)在诊断精度上,相比于 4 种常用的 MKL 算法,本文模型在实现低漏警率、低虚警率的同时能显著提升诊断的正确率。对于旋转变压器激励发生电路,M_1-LCMKELM 将诊断精度平均提升了 3.80%,而 M_2-LCMKELM 将诊断精度平均提升了 5.98%。

2)在诊断时间上,相比于 2 种流行的 LMKL 算法,本文模型具有相似、甚至更短的训练时间;同时得益于 ELM 快速的计算效率,可以实现更短的测试时间。

3)M_1-LCMKELM 产生局部权重的稀疏解,M_2-LCMKELM 产生局部权重的非稀疏解。这使得 M_2-LCMKELM 的诊断精度要优于 M_1-LCMKELM,但是 M_1-LCMKELM 的测试时间要优于 M_2-LCMKELM。

参考文献(References)

[1] LUO H, WANG Y R, LIN H, et al. Module level fault diagnosis for analog circuits based on system identification and genetic algorithm[J]. Measurement, 2012, 45(4): 769-777.

[2] 孙伟超,李文海,李文峰. 融合粗糙集与 D-S 证据理论的航空装备故障诊断[J]. 北京航空航天大学学报, 2015, 41(10): 1902-1909.
SUN W C, LI W H, LI W F. Avionic devices fault diagnosis based on fusion method of rough set and D-S theory[J]. Journal of Beijing University of Aeronautics and Astronautics, 2015, 41(10): 1902-1909(in Chinese).

[3] KNÜPPEL T, BLANKE M, ØSTERGAARD J. Fault diagnosis for electrical distribution systems using structural analysis[J]. International Journal of Robust and Nonlinear Control, 2014, 24(8-9): 1446-1465.

[4] JAMIL T, MOHAMMED I. Simulation of VICTOR algorithm for fault diagnosis of digital circuits[J]. International Journal of Computer Theory and Engineering, 2015, 7(2): 103-107.

[5] DAI X W, GAO Z W. From model, signal to knowledge: A data-driven perspective of fault detection and diagnosis[J]. IEEE Transactions on Industrial Informatics, 2013, 9(4): 2226-2238.

[6] 蒋栋年,李炜. 基于自适应阈值的粒子滤波非线性系统故障诊断[J]. 北京航空航天大学学报, 2016, 42(10): 2099-2106.
JIANG D N, LI W. Fault diagnosis of particle filter nonlinear systems based on adaptive threshold[J]. Journal of Beijing University of Aeronautics and Astronautics, 2016, 42(10): 2099-2106(in Chinese).

[7] GAO Z W, CECATI C, DING S X. A survey of fault diagnosis and fault tolerant techniques—Part I: Fault diagnosis with model-based and signal-based approaches[J]. IEEE Transactions on Industrial Electronics, 2015, 62(6): 3757-3767.

[8] SAHRI Z B, YUSOF R B. Support vector machine-based fault diagnosis of power transformer using k nearest-neighbor imputed DGA dataset[J]. Journal of Computer and Communications, 2014, 2(9): 22-31.

[9] YIN G, ZHANG Y T, LI Z N, et al. Online fault diagnosis method based on incremental support vector data description and extreme learning machine with incremental output structure[J]. Neurocomputing, 2014, 128: 224-231.

[10] GÖNEN M, ALPAYDIN E. Multiple kernel learning algorithms[J]. Journal of Machine Learning Research, 2011, 12: 2211-2268.

[11] YE F M, ZHANG Z B, CHAKRABARTY K, et al. Board-level functional fault diagnosis using multikernel support vector machines and incremental learning[J]. IEEE Transactions on Computer-aided Design of Integrated Circuits and Systems, 2014, 33(2): 279-290.

[12] LI Y X, REN C Q, BO J Y, et al. The application of GMKL algorithm to fault diagnosis of local area network[J]. Journal of Networks, 2014, 9(3): 747-753.

[13] RAKOTOMAMONJY A, BACH F R, CANU S, et al. SimpleMKL[J]. Journal of Machine Learning Research, 2008, 9: 2491-2521.

[14] HAN Y N, YANG K, MA Y L, et al. Localized multiple kernel learning via sample-wise alternating optimization[J]. IEEE Transactions on Cybernetics, 2014, 44(1): 137-147.

[15] SONG Y, ZHENG Y T, TANG S, et al. Localized multiple kernel learning for realistic human action recognition in videos[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2011, 21(9): 1193-1202.

[16] GÖNEN M, ALPAYDIN E. Localized algorithms for multiple kernel learning[J]. Pattern Recognition, 2013, 46(3): 795-

http://bhxb.buaa.edu.cn

807.

[17] WANG X M, HUANG Z X, DU Y J. Improving localized multiple kernel learning via radius-margin bound[J]. Mathematical Problems in Engineering, 2017, 2017: 4579214.

[18] HAN Y N, LIU G Z. Probability-confidence-kernel-based localized multiple kernel learning with L_p norm[J]. IEEE Transactions on Systems, Man and Cybernetics-Part B: Cybernetics, 2012, 42(3): 827-837.

[19] HAN Y N, YANG K D, LIU G Z. L_p norm localized multiple kernel learning via semi-definite programming[J]. IEEE Signal Processing Letters, 2012, 19(10): 688-691.

[20] FREY B J, DUECK D. Clustering by passing messages between data points[J]. Science, 2007, 315(5814): 972-976.

[21] NAPOLEON D, BASKAR G, PAVALAKODI S. An efficient clustering technique for message passing between data points using affinity propagation[J]. International Journal on Computer Science and Engineering, 2011, 3(1): 8-13.

[22] SOKOLOVA M, LAPALME G. A systematic analysis of performance measures for classification tasks[J]. Information Processing and Management, 2009, 45(4): 427-437.

[23] PHOUNGPHOL P, ZHANG Y Q, ZHAO Y C. Robust multiclass classification for learning from imbalanced biomedical data[J]. Tsinghua Science and Technology, 2012, 17(6): 619-628.

作者简介：

张伟 男，博士研究生。主要研究方向：机载电子设备状态监测与故障诊断。

许爱强 男，博士，教授，博士生导师。主要研究方向：复杂电子系统自动测试与诊断技术。

Localized multi-kernel diagnosis model for avionics based on affinity propagation clustering

ZHANG Wei[1], XU Aiqiang[1,*], PING Dianfa[1], XIA Fei[2]

(1. Naval Aeronautical University, Yantai 264001, China; 2. Information and Communication Branch Office, State Grid Liaoyang Electric Power Supply Company, Liaoyang 111000, China)

Abstract: In consideration of the low diagnosis accuracy for avionics functional module fault, a new off-line localized clustering multi-kernel extreme learning machine (LCMKELM) diagnosis model is proposed in this paper by combining the capabilities of multi-resolution interpretation and local feature self-adaptive representation from localized multi-kernel learning (LMKL) with the characteristic of high-performance operation from extreme learning machine (ELM). In order to avoid overfitting issue, affinity propagation (AP) clustering is used to make full use of the underlying localities in the training data and effectively reduce the computational complexity. Considering that the updating of localized kernel weights in dual optimization form of kernel ELM (KELM) is a difficult quadratic nonconvex problem, gating function M_1 and M_2 are respectively constructed to approximate localized weights by analyzing the clustering characteristics in input space and feature space. The proposed method is applied to actual fault diagnosis task of rotary transformer excitation generating circuit, and the experimental results show that the proposed method has the lower false alarm rate and missing alarm rate in comparison with four state-of-the-art multi-kernel learning algorithms, and meanwhile the diagnosis accuracy is averagely increased by 3.80% when M_1 gating model is used, and increased by 5.98% when M_2 gating model is used. Moreover, compared with canonical LMKL algorithms, the proposed method obtains similar training time cost, but it has less testing time cost.

Keywords: fault diagnosis; multi-kernel learning; alternating optimization; affinity propagation (AP); gating model; localized algorithm

Received: 2017-10-16; **Accepted**: 2017-11-17; **Published online**: 2018-01-25 11:04

URL: kns.cnki.net/kcms/detail/11.2625.V.20180123.1858.003.html

Foundation items: National Natural Science Foundation of China (61571454); Natural Science Foundation of Shandong Province, China (ZR2016FQ03)

* **Corresponding author**. E-mail: hjhyautotest@sina.com

http://bhxb.buaa.edu.cn