

结合 DCT 和 SVM 的故障诊断方法研究

舒云星^{1,2}, 张焕龙²

SHU Yun-xing^{1,2}, ZHANG Huan-long²

1. 武汉理工大学 机电工程学院, 武汉 430070

2. 洛阳理工学院 计算机与信息工程系, 河南 洛阳 471003

1. School of Mechatronic Engineering, Wuhan University of Technology, Wuhan 430070, China

2. Department of Computer and Information Engineering, Luoyang Institute of Science and Technology, Luoyang, Henan 471003, China

E-mail: zhl888zhl_2@163.com

SHU Yun-xing, ZHANG Huan-long. Study on fault diagnose method combining DCT and SVM. *Computer Engineering and Applications*, 2008, 44(35): 200-202.

Abstract: Aiming at the problem of too high dimension of sample data leading to too much SVM studies in the researches of fault diagnose, This paper proposes a new SVM fault diagnose method based on DCT, using the "energy compaction" and "high frequency suppression" features of DCT, which used in the process of image compression. At first, reduce the dimension of fault samples in light of the method of DCT. Secondly, train the main coefficients of DCT instead of the original samples using SVM so as to restrain the effect of noise for fault classification greatly, and reduce the computation in the process of diagnosis. Finally, the validity of the algorithm is proved by experiment simulation.

Key words: fault diagnose; Support Vector Machine(SVM); Discrete Cosine Transform(DCT); high frequency suppression

摘要: 针对故障诊断研究中, 样本数据维数过高导致故障模式分类时 SVM 学习强度太大的问题, 利用 DCT 方法在降噪处理时体现出“能量集中”和“高频抑制”的特性, 提出一种基于 DCT 的 SVM 故障诊断方法。先对故障样本进行 DCT 降维, 再利用 SVM 方法对主要维离散余弦系数进行模式训练来代替对故障样本的直接训练, 从而大大抑制了噪声对故障分类的影响, 同时也减少了诊断运算量, 最后通过实验仿真验证了算法的有效性。

关键词: 故障诊断; 支持向量机(SVM); 离散余弦变换(DCT); 高频抑制

DOI: 10.3778/j.issn.1002-8331.2008.35.060 **文章编号:** 1002-8331(2008)35-0200-03 **文献标识码:** A **中图分类号:** TP277

1 引言

故障诊断已成为现代智能控制系统中一个重要的环节, 但其因为故障样本维数过高、故障与征兆间关系的不确定性和故障样本不断更新等因素的影响, 使得故障诊断的研究出现很多困难。而一般故障诊断的方法可分为故障样本的预处理和模式分类两部分, 前者主要针对样本维数过高, 数据之间冗余过多的问题, 通过算法处理将对分类贡献较大的样本属性保留, 其余忽略不计从而减少模式分类时运算量过大的问题, 也有针对训练样本学习强度太大的问题, 对故障样本经过算法处理进行粗分类后再进行模式分类, 这也大大缩短了分类时间, 但其对故障样本数据有一定的局限性, 通用性不高; 而后者主要是实现故障样本模式的学习、分类及预测。

对于样本数据降维的研究, 常见的方法主要有粗糙集属性约简^[1], 基于 PCA 的特征降维^[2], 小波故障特征提取等。而粗糙集理论是通过寻找样本各个属性知识对决策集合的支持程度以确定核心知识, 从而实现对无关知识的属性进行约简, 故相

对计算量较大; PCA 降维从故障数据样本组成的协方差矩阵入手, 通过对其特征值累计方差贡献率大小的分析确定模式分类主元, 最终实现特征降维, 但其完全依赖于原始样本数据的变化, 故对噪声干扰十分敏感; 小波方法将故障样本分频率实现特征提取, 其需要较大的样本数量, 计算也相对复杂。故引入离散余弦变换来实现故障样本数据的降维处理, 其主要优点在于“能量聚集性”, 可以十分方便地确定特征主要属性; 又因为一般主要特征数据集往往被包含在中、低频区域, 故可以利用其“高频抑制性”实现故障样本的特征降维, 从而提高抗噪能力, 使算法对样本中出现少量异常数据时也能有很好的适应性。

对于模式分类方法的研究, 主要方法有神经网络法^[3]和支持向量机^[4]等。而神经网络学习算法存在收敛速度缓慢、容易陷入局部极小以及网络结构难以确定等缺陷, 以至于大大影响网络的学习效率和分类精度, 更重要的是神经网络需要一个大量的故障样本用来训练, 但一般故障样本均比较小, 这也极大地限制了其在故障诊断中的应用, 因此很多研究者都致力于基于

基金项目: 河南省科技厅科技攻关项目资助(No.0624440059)。

作者简介: 舒云星(1962-), 男, 教授, 博士, 主研方向: 智能算法研究和故障诊断等; 张焕龙(1981-), 男, 助教, 主研方向: 图像处理和模式识别。

收稿日期: 2008-09-10 修回日期: 2008-10-13

支持向量机的故障诊断方法研究。支持向量机(Support Vector Machine, SVM)基于结构风险最小化原则,表现出了很强的泛化能力,可以很好地克服维数灾难和过拟合等传统算法所不可规避的问题,并较好地实现了对线性、非线性样本数据的学习和分类。

针对故障样本维数过高的问题,将离散余弦变换方法引入到故障诊断研究中,利用 DCT“能量聚集性”和“高频抑制性”将故障样本维数降低,再通过 SVM 方法实现对故障的准确分类,并使算法具有一定的抗噪声能力。

2 基于 DCT 的特征降维

离散余弦变换(Discrete Cosine Transform, DCT)是一种与傅里叶变换紧密相关的数学运算,在傅里叶级数展开式中,如果展开的函数是实偶函数,那么其傅里叶级数中只包含余弦项,再将其离散化可导出余弦变换,称之为离散余弦变换,它可实现信号空频域间的相互转化,通过处理转化后的余弦系数,利用主要的系数重构逼近原始信号从而获取与原数据误差最小的特征维数。

2.1 DCT 基本原理

离散余弦变换是一种常见的正交变换,一维离散变换的定义如下^[5]:

$$F(u)=C(u)\sqrt{\frac{2}{N}}\sum_{t=0}^{N-1}f(t)\cos\left[\frac{\pi u(2t+1)}{2N}\right] \quad (1)$$

其中 $F(u)$ 为第 u 个余弦变换系数, u 为广义频率变量, $u=0, 2, \dots, N-1$; $f(t)$ 为实域中 N 个离散的点序列, $t=0, 1, 2, \dots, N-1$ 。而一维离散反余弦变化可表示如下:

$$f(t)=C(u)\sqrt{\frac{2}{N}}\sum_{u=0}^{N-1}F(u)\cos\left[\frac{\pi u(2t+1)}{2N}\right] \quad (2)$$

而式中:

$$C(u)=\begin{cases} 1/\sqrt{2} & u=0 \\ 1 & u\neq 0 \end{cases} \quad (3)$$

故障样本经过离散余弦变换后,将在频率域中进行操作,分解后的信号包含直流分量和交流分量, $u=0$ 时获得的余弦系数是直流分量,其代表着信号中变化比较平缓的数据即低频信息,它可以基本体现原信号的主要特征,其它均为交流分量对原信号特征的体现影响较小,故可以将 u 较大的那部分系数置零,然后重构出十分逼近原信号的近似信号。另一方面,随着频率 u 的增大相当于高频信息丰富,当舍弃较大 u 值实现信号重构时,相当于将高频的信息减弱,而噪声干扰信息一般为高频信息,这样用近似信号训练时就避免了奇异样本数据的干扰。

2.2 基于 DCT 的特征降维

假设现存在一组故障样本为 $x_i, i=1, 2, \dots, m$, 而故障样本具有 n 个特征属性,也就是 x_i 为 n 维的向量,这样训练样本可以构成一个 $m \times n$ 的矩阵 A , 其中 m 代表训练样本的个数,而 n 代表样本的维数。

首先利用 DCT 算法对 x_i 进行一维变换获取离散余弦系数矩阵 D_i , x_i 和 D_i 均为一个 n 维的行向量, x_i 以矢量的形式排列构成的矩阵表示故障样本的所有数据,而 D_i 相应构成的系数矩阵则表示离散余弦变换后原样本数据的特征描述,如果单纯使用系数矩阵作为原样本矩阵的特征描述,因为系数矩阵的阶次和原矩阵的维数相同,所以图像特征矩阵的维数仍然很高,并

没有达到降维的作用。但是, DCT 系数矩阵中很多元素为零或者近似为零,而且 u 值较大时, DCT 的系数值 D_i 很小;而数值较大的 D_i 主要分布在 u 值较小区域,即前若干维的数据就可以极大程度地重构出样本数据的主要特征。采用区域模板 P 对 D_i 点乘,提取系数矩阵中较大的系数,并将提取的系数以矢量形式排列起来,实现特征降维。描述如下:

$$H_i = P \times D_i \quad i=1, 2, \dots, n \quad (4)$$

$$P = \begin{bmatrix} 1 & \dots & 1 & 0 & \dots & 0 \\ 1 & \dots & 1 & 0 & \dots & 0 \\ 1 & \dots & 1 & 0 & \dots & 0 \\ \vdots & \dots & 1 & \vdots & \dots & \vdots \\ 1 & \dots & 1 & 0 & \dots & 0 \\ 1 & \dots & 1 & 0 & \dots & 0 \end{bmatrix} \quad (5)$$

区域模板 P 除了前 s 维元素外,其余均被置零, s 的大小决定了最终样本的维数,则:

$$H_i = [h_{i1}, h_{i2}, \dots, h_{is}, 0, 0, \dots, 0] \quad (6)$$

其中 $h_{is} \neq 0, (s=1, 2, \dots, n)$, 由于零值对故障样本的特征来说没有意义,所以只取中非零值的元素,即 $\bar{H}_i = [h_{i1} \ h_{i2} \ \dots \ h_{is}]$ 。按照从上到下的顺序将 \bar{H}_i 排列起来组成特征矩阵 H , 其大小为 $m \times s$, 相对于大小为 $m \times n$ 的原特征矩阵维数大大下降。

$$H = \begin{bmatrix} H_1 \\ \vdots \\ H_m \end{bmatrix}_{m \times s} = \begin{bmatrix} h_{11} & \dots & h_{1s} \\ \vdots & & \vdots \\ h_{m1} & \dots & h_{ms} \end{bmatrix}_{m \times s} \quad (7)$$

离散余弦变换后信号的主要能量集中在矩阵 H 的若干维上,频率越高即余弦系数越小代表的信息量越小对原信号的影响程度也越小,故可以利用能够较大体现原信号成分的较少维余弦变化系数作为故障样本的特征维进行训练,从而实现特征降维。同时从以上过程也可以看到,将高频位置零使得原信号中的高频信息大大减弱,而通过对主要余弦系数进行训练代替原样本数据进行模式训练,很明显提高了对噪声干扰的抵抗能力。

3 基于 DCT—SVM 的故障诊断方法

3.1 支持向量机

支持向量机能较好地解决小样本、非线性和局部最小点等实际问题,主要思想是:寻找最优分类超平面使得训练样本中的两类样本点能被准确分类,并且类间距保持最大;对于线性不可分问题,通过核函数将低维输入空间的数据影响到高维空间从而将不可分问题在高维中转化成可分问题,然后在新空间中寻求最优超球面。

设有 n 个输入故障模式为 x_i , 其属于的类别为 y_i , 可表示为 $\{(x_i, y_i)\}$ 。其中 $y \in \{-1, 1\}, i=1, \dots, n; n$ 为输入样本的维数,具体分类描述如下:

(1)在线性可分的情况下,就会存在一个超平面可以描述为: $w \cdot x + b = 0$, 为使分类面对所有样本正确分类且具有分类间隔,要求满足如下约束:

$$y_i [x_i \cdot w + b] - 1 \geq 0, i=1, 2, \dots, n \quad (8)$$

目标是分类间隔 $\rho = 2 / \|w\|$ 取得最大值。使式(8)等号成立的样本叫做支持向量,然后用支持向量来获得分类最优超平面从而实现对未知故障样本的正确分类。

(2)在线性不可分的情况下,可以在式(8)中增加一个松弛项 $\xi_i \geq 0$, 成为:

$$y_i[(w \cdot x_i) + b] - 1 - \xi_i \geq 0, \quad i=1, 2, \dots, n \quad (9)$$

将目标改为求:

$$\rho = 1/2 \|w\|^2 + C \left(\sum_{i=1}^n \xi_i \right) \quad (10)$$

的值最小,且 $0 \leq \alpha_i \leq C$ 其中, C 是惩罚因子,即综合考虑最少错分样本和最大分类间隔,获得广义最优分类面,从而实现对未知故障样本在错分最少的情况下正确分类。

3.2 DCT—SVM 故障诊断算法描述

设故障样本数据构成一个 $m \times n$ 的矩阵 A , 其可以用列向量的形式表示为 $[x_1, x_2, \dots, x_m]$, 每一个向量表示为 $x_i = [y_1, y_2, \dots, y_n]$, 其中 m 代表训练样本的个数,而 n 代表样本的维数。

(1)假设一维 DCT 算子为 D , 对每一个向量实现 $D(x_1), D(x_2), \dots, D(x_m)$ 离散余弦变换获得系数矩阵 $B_{m \times n}$ 。

(2)将离散余弦系数矩阵 $B_{m \times n}$ 中近似为零的系数置为零,然后将为零的系数去掉,系数矩阵变为近似矩阵 $C_{m \times s}$, $s(1 \leq s \leq n)$ 为新的维数。

(3)采用支持向量机方法对 $C_{m \times s}$ 进行故障模式的训练,从而代替了原始样本 $B_{m \times n}$ 的训练,实现了海量数据训练时样本维数的下降。

(4)将测试样本做上述的处理,将降维处理后测试样本的离散余弦系数代替原始测试样本进行诊断,获取诊断正确率。如果正确率太低说明维数减少太多,数据特征淹没太多,重新转到第(2)步,进行确定新的 s 值,直到获得最高的诊断正确率。

4 实验仿真

实验以现有的故障数据 1 为例^[6], 样本为通风机的 5 种常见故障为例,每一种故障有 8 个主要特征,训练样本集为 15 个 8 维的向量组成;故障数据 2 采用 vehicle2 故障数据集,其故障模式有 4 类,每一种故障有 18 个主要特征,训练样本集为 846 个 18 维的向量组成。实验分别采用直接 SVM 分类、DCT 降维后再采用 SVM 分类和粗糙集降维后再采用 SVM 三种方法实现故障分类;接着对故障样本中加入一些奇异值作为噪声影响,再进行同样三次分类实验,来体现所提方法的优势所在,采用支持向量机中 one-vs-one 方法进行多类模式分类,最后对实验数据比较分析,验证本文方法的有效性。仿真结果如表 1 和表 2 所示。

表 1 故障诊断方法比较

分类方法	原始样本大小	特征样本大小	训练样本维数	分类正确率/(%)
数据 1				
直接 SVM 分类	12×8	12×8	8	100
DCT-SVM 分类	12×8	12×3	3	100
粗糙集 SVM	12×8	12×5	5	100
数据 2				
直接 SVM 分类	846×18	846×18	18	95.4
DCT-SVM 分类	846×18	846×8	8	96.2
粗糙集 SVM	846×18	846×11	11	95

表 1 记录了两组故障样本分类方法的实验结果,可以看出直接方法、粗糙集方法和本文方法对数据 1 均可以实现正确分类,而粗糙集方法和 DCT 方法均在训练数据的维数上有明显的降低,这样就大大减小了模式分类时的运算量,但就这两者比较而言本文方法在特征降维方面又稍显优势;对于数据 2 同

表 2 加噪后故障诊断方法比较

分类方法	原始样本大小	特征样本大小	训练样本维数	分类正确率/(%)
数据 1				
直接 SVM 分类	12×8	12×8	8	66.6
DCT-SVM 分类	12×8	12×5	5	100
粗糙集 SVM	12×8	12×6	6	100
数据 2				
直接 SVM 分类	846×18	846×18	18	92.5
DCT-SVM 分类	846×18	846×12	12	94.2
粗糙集 SVM	846×18	846×14	14	93.3

样可以看出本文方法在取得较好降维效果的同时,也保证了诊断准确率。

表 2 记录了对两组故障样本分别加噪后再分类的实验结果,实验加入椒盐噪声(采用 Matlab 中的加噪方法,参数值为 0.2)。从数据 1 结果可以看出直接分类方法准确率已经变成 66.6%,而粗糙集和本文方法仍然可以准确实现分类,只是维数相对于无噪声时均增大了,运算量也随之稍有增大,两者对噪声都有一定的适应性,然而本文方法在特征降维方面有更好的效果;对于数据 2 也再次说明了上述的结论。

通过表 1 和表 2 显示的结果可以看出:

(1)在故障诊断预处理中,DCT 利用其“能量集中性”进行特征降维,取得了好的效果,降低了故障样本数据的维数,同时用特征余弦系数代替原始数据进行训练和分类,大大降低了算法实现的运算量。

(2)故障数据加噪声后,依据 DCT 的“高频抑制性”在特征降维时将具有高频特征的故障数据削弱,并使特征数据集尽可能大地逼近原始故障数据集,故在牺牲降低数据维数的情况下对噪声影响有很大的适应性。

5 结论

通过理论和实验分析了 DCT 利用“能量集中”和“高频抑制”特性对故障样本进行特征降维的可行性,并且将其和 SVM 相结合应用于两组故障数据的实验中。通过故障样本数据在正常和加噪两种情况下采用直接 SVM 分类、粗糙集分类和 DCT 分类 3 种方法的实验,充分说明了经过 DCT 降维后结合 SVM 实现故障诊断时,不仅可以降低故障样本数据的维数,减少训练、学习和分类的运算量,而且对故障样本中出现的少量奇异数据有很好的适应性。

参考文献:

- [1] 郭小荃,马小平.基于粗糙集的故障诊断特征提取[J].计算机工程与应用,2007,43(1):221-224.
- [2] Xu Xin,Wang Xue-ning.An adaptive network intrusion detection method based on PCA and support vector machines[C]//LNCS 3584,2005:696-703.
- [3] 刘安,刘春生.基于 RBF 神经网络的非线性系统故障诊断[J].计算机仿真,2007,24(2):141-144.
- [4] Gunn S.Support vector machines for classification and regression[R].UK:University of Southampton,1998.
- [5] 胡永刚,吴翔,王洪志,等.高维数据降维的 DCT 变换[J].计算机工程与应用,2006,42(32):21-23.
- [6] 荆双喜,华伟.基于小波-支持向量机的矿用通风机故障诊断[J].煤炭学报,2007,32(1):98-102.