

# 一种基于关联频繁模式的振动数据流挖掘框架

张艳梅<sup>1</sup> 陆 伟<sup>2</sup> 杨余旺<sup>2</sup>

(1. 广东培正学院电子商务系, 广州, 510830; 2. 南京理工大学计算机科学与工程学院, 南京, 210094)

**摘 要:** 针对大型滚转机器轴承故障诊断应用场景中传统故障识别技术通常存在诊断识别精度低的问题, 在频域分析基础上提出了一种新的数据挖掘框架——关联频繁模式集挖掘框架 (Associated frequency patterns mining framework, AFPMF), 由数据预处理、关联频繁模式集挖掘和故障状态监测组成。首先, 在数据预处理过程中, AFPMF 在时域上使用时间窗分块划分机械振动数据流, 再使用傅立叶变换对数据流进行时频变换实现故障频率特征提取。其次, 使用基于滑动窗的关联频繁模式树构建压缩树, 求解关联频繁模式集, 实现数据挖掘过程。最后, 根据数据挖掘结果中出现的振动频率判别潜在故障, 从而实现监测故障状态。通过对比 AFPMF 和传统方法在轴承故障诊断应用场景的实验结果可知, 相比传统方案, AFPMF 具有更优的故障识别性能。

**关键词:** 机器状态监测; 轴承故障; 关联频繁模式树; 振动数据流

**中图分类号:** TP391.5      **文献标志码:** A

## Novel Data Mining Framework for Vibration Data Stream Based on Associated Frequency Patterns

Zhang Yanmei<sup>1</sup>, Lu Wei<sup>2</sup>, Yang Yuwang<sup>2</sup>

(1. Department of Electronic Commerce, Guangdong PeiZheng College, Guangzhou, 510830, China; 2. School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, 210094, China)

**Abstract:** In the scenarios of diagnosing bearing faults for large rotary machinery, the traditional fault identification technique usually has the problem of low identification accuracy. Hence, based on the frequency-domain analysis, a novel data mining framework of frequency patterns mining framework (AFPMF) is proposed in this paper, which consists of data pre-processing, associated frequency pattern mining process and fault status monitoring. In the data pre-processing of AFPMF, the time window is adopted to divide the machinery vibration data stream into multiple sub-blocks, and then fast Fourier transform (FFT) is employed to make the data sub-blocks time-frequency transform for frequency feature extraction. The associated frequency pattern tree with sliding window is also used to build a compact tree for data mining. Finally, the potential fault is identified according to the vibration frequency existing in the mining results. Thus the fault status monitoring is realized. The comparison results of AFPMF and the traditional methods in the bearing fault diagnosis show that AFPMF has higher identification accuracy than other traditional ones.

**基金项目:** 国家自然科学基金(61640020)资助项目。

**收稿日期:** 2017-12-14; **修订日期:** 2018-03-07

**Key words:** machine condition monitoring; bearing fault; associated frequent pattern tree; vibration data stream

## 引言

轴承是大多数滚转机器的重要组件,约40%的滚转机器故障是由轴承故障造成,主要包括外圈、内圈和滚珠轴承故障<sup>[1]</sup>。目前,机器状态监测(Machine condition monitoring, MCM)技术主要是对磨损产物、电机电流、噪声、气温和振动信号等进行监测和分析,其中振动信号分析凭借其可靠性和灵敏性成为最有效的监测对象<sup>[2]</sup>。不同的故障类型会引起不同的振动频率成分,主要和轴承几何形状及运行速度有关系,其外圈、内圈和滚珠轴承故障所对应的振动频率成分函数表达式分别为<sup>[3]</sup>

$$f_{of} = \frac{n}{2} f_{rm} \left( 1 - \frac{d_{ball}}{d_{pitch}} \cos\phi \right) \quad (1)$$

$$f_{if} = \frac{n}{2} f_{rm} \left( 1 + \frac{d_{ball}}{d_{pitch}} \cos\phi \right) \quad (2)$$

$$f_{bf} = \frac{d_{pitch}}{2d_{ball}} f_{rm} \left( 1 - \left( \frac{d_{ball}}{d_{pitch}} \right)^2 \cos\phi^2 \right) \quad (3)$$

式中: $f_{rm}$ ,  $d_{pitch}$ ,  $d_{ball}$ ,  $n$ 和 $\phi$ 分别为旋转频率、节圆直径、滚珠直径、滚珠数目和接触角。振动分析技术包括时域、频域和时频域分析。时域分析难以分离出不同振动源,而由于振动频率特性,频域分析成为主要分析手段,其中以能够将时频域振动信号转换为离散频率成分的快速傅立叶变换(Fast Fourier transform, FFT)为主<sup>[4]</sup>。因此,本文使用FFT对振动信号作频域分析。

信号频率域分析主要是从机器振动数据中提取故障信号特征,并使用机器学习技术从正常信号中识别出故障特征<sup>[5]</sup>。由于机器故障类型多和故障发生的偶然性使故障收集变得很难,通常使用模式分类技术解决此问题<sup>[6-9]</sup>,涉及小波技术、人工神经网络和支持向量机技术,但均存在故障识别精确度低的问题。目前已有研究将数据挖掘技术引入到故障诊断研究中,通过使用粗糙集、决策树融合等实现旋转机械、感应电机等的故障诊断决策<sup>[10-11]</sup>,相比传统模式分类技术具有较优的识别性能,结果显示数据挖掘技术在故障诊断领域具有很大的发展潜力。

频繁模式挖掘是数据挖掘领域的一种基本方法,但由于数据流环境具有数据量大、实时性、数据预知性差等特点,传统频繁模式挖掘算法难以适应。关联频繁模式挖掘算法由于具有更高的挖掘效率更适用于数据流环境,主要分为广度优先算法和深度优先算法两大类,广度优先算法有Apriori, Apriori-Hybrid等<sup>[12]</sup>,其在密集数据库环境下效率比较低。深度优先算法则以FP-growth<sup>[13]</sup>, FP-Streams<sup>[14]</sup>为代表,其中FP-Streams算法使用频繁模式树存储过去时间窗口中的频繁模式项集信息,但当事务长度增加时算法效率会降低。由于现实应用中,人们更关心当前时间窗口内到达的数据流信息,因此一些学者将滑动窗口引入到数据挖掘之中。文献[15]研究了时间窗口长度固定和变化条件下的频繁模式挖掘算法,文献[16]针对滑动窗口数据挖掘算法,提出了一种数据结构闭合枚举树,用来存储动态选定项及其边界,降低了算法时间代价。因此,本文将关联频繁模式挖掘中的深度优先算法引入到对轴承振动数据流的数据挖掘中,再借鉴上述算法使用的滑动时间窗口思想分块求取频繁模式集。

本文提出了一种新的数据挖掘框架——关联频繁模式集挖掘框架(Associated frequency patterns mining framework, AFPMF),设计了一种基于滑动窗口的关联频繁模式树(Sliding window associated frequency pattern-tree, SWAFP-T)结构和SWAFP算法,挖掘振动数据流得到关联频繁模式集,识别故

障频率,获得优于现有技术的故障识别性能。频繁模式挖掘技术受最小支持阈值约束,阈值大小影响频繁模式集数目。由于轴承振动信号从各角度分析会产生大量数据,需要寻找合适的筛选方法在振动数据中找到具有强相关性的频率特征模式。因此,本文将使用关联频繁模式从振动数据中寻找具有强相关性的频率特征模式,对通过已知故障数据得到的故障特征进行FFT分析,以提取故障的频域成分信息。最后,方案根据频繁模式集对轴承的缺陷状态实施监测,进而避免了轴承故障的发生。实验验证了方案有效性,结果表明本文所提出方案具有优越的检测效率和识别精确度。

## 1 数据预处理

数据挖掘框架 AFPMF 主要由数据预处理、关联频繁模式挖掘和故障频率识别构成。数据预处理过程主要处理从轴承数据寄存器中收集的振动数据。首先使用大小合适的矩形窗将振动数据划分为多个具有相同时间大小的子窗,长度为  $l$  的信号  $X$  被采样点数  $W$  的矩形窗划分为  $m$  个时间子窗,表达式为

$$m = \frac{l}{W} \quad (4)$$

其中,振动信号  $X$  的任意时间子窗  $i$  表示为

$$x_i = X_i(\omega) \quad (5)$$

式中:  $i = 1, 2, 3, \dots; \omega = 1, 2, 3, \dots, W$ 。受机器运行状态和环境影响,振动数据包含的故障特征信号会受不同程度水平噪声影响<sup>[17]</sup>。为分析不同噪声水平影响,时间子窗  $x_i$  被高斯白噪声影响,表达式为

$$x_i = X_i(\omega) + \text{AWGN}(-dB) \quad (6)$$

由于时间域信号的时变特性,将时间子窗信号  $x_i$  转换到频域,其频域  $f_i$  表达式为

$$f_i = FT(x_i) \quad (7)$$

其中当窗口尺寸为  $W$  时,  $f_i$  包含  $W/2 + 1$  个频率,对  $f_i$  归一化,即

$$\bar{f}_i = \frac{f_i}{\max(f_i)} \quad (8)$$

其中频率具有连续幅值,关联模式挖掘要求频率使用二值表示即为激活态(1)或者非激活态(0),使用阈值  $\varepsilon$  将  $\bar{f}_i$  中的每个频率幅值转换为二值,其表达式为

$$S_i(j) = \begin{cases} 1 & \bar{f}_i(j) \geq \varepsilon \\ 0 & \bar{f}_i(j) < \varepsilon \end{cases} \quad (9)$$

式中:  $j = 1, 2, 3, \dots, \frac{W}{2} + 1; i = 1, 2, 3, \dots, m$ 。此外,使用数据集  $T$  表示所有  $m$  个时间子窗包含的频率信息,即

$$T = T_1, T_2, T_3, \dots, T_m \quad (10)$$

$$T_i = j \mid (S_i(j) = 1) \quad (11)$$

式中:  $T_i$  包括在时间子窗  $i$  内所有幅值为 1 的频率信息。

## 2 挖掘关联频繁模式集

### 2.1 基础概念

设  $F = \{f_1, f_2, \dots, f_n\}$  为故障振动数据预处理后的频率集。假设时间被划分为等长时隙  $t =$

$\{t_1, t_2, \dots, t_q\}, t_{j+1} - t_j = \lambda$  和  $j \in [1, q - 1], \lambda$  为时隙大小。集合  $P = \{f_1, f_2, \dots, f_p\} \subseteq F$  为频繁模式, 振动数据流被定义为  $VDS = [e_1, e_2, \dots, e_n], r \in [1, n], e_r$  表示元组  $E(E_{ts}, P)$ , 其中  $P$  为元组所在相同时间  $E_{ts}$  子窗内的频繁模式, 设  $size(E)$  为  $E$  大小, 表示频率数目。设每个窗口  $W$  内包含长度相等的非重叠序列的批, 设有  $M$  序列和  $N$  批, 批由  $M/N$  构成, 故批大小为  $\lfloor M/N \rfloor$ , 设定窗以批为单位进行串序滑动。图 1 给出了振动数据流结构示意图。

**定义 1** 窗口  $W$  中模式  $P$  支持集为  $SUP_w(P)$ , 如果某模式支持度大于支持度阈值  $MinSUP$ , 有  $0 \leq MinSUP \leq |W|$ , 则认其为一个频繁模式。再者, 模式  $P$  的最大频率支持集为  $MaxfreqSUP_w(P) = \max(SUP(f_j) | \forall f_j \in P)$ 。

**定义 2** 设窗口  $W$  中模式  $P$  信任度为  $\alpha_w(P)$ , 其表达式为

$$\alpha_w(P) = \frac{SUP_w(P)}{MaxfreqSUP_w(P)} \quad (12)$$

**定义 3** 如果  $\alpha_w(P)$  大于或等于窗口  $W$  中给定的最小信任度阈值  $MinallConf$ , 则称其为关联频繁模式。

当给定振动数据流  $VDS, |W|, MinSUP$  和  $MinallConf$  等数据时, 将关联频繁模式挖掘问题转换为  $|W|$  中大于某阈值条件下的模式集合寻找问题。

## 2.2 树构建过程

数据流  $VDS$  由无限长序列构成, 批由非空集的序列构成。SWAFP-T 树构建过程分为插入阶段和重构压缩阶段。

插入阶段, SWAFP-T 树根据频率在数据流中出现的顺序依次安排频率顺序, 并通过插入的方式排列  $VDS$  中先后出现的每个序列, 并且维护一个存储所有频率支持度的频率头列表  $F$ -list, 完成  $SWAFP-T_A$  初始化。

初始化阶段  $SWAFP-T_A$  结束后, 开始重构压缩阶段, 通过本阶段得到一个高度压缩的  $SWAFP-T$ , 能够占用更小的内存空间和更快的挖掘处理速度。重构压缩处理过程使用文献[18]中提出的分支排序方法 (Branch sorting method, BSM)。BSM 使用合并排序排列前缀树上的每条路径, 该方法首先移除未排序路径然后给其他路径排序, 最后再将之前移除的路径重新插入排列。此外, 使用文献[19]提出的压缩方法选择树上每条分支具有相同支持度的频率节点并将其合并为一个节点。SWAFP-T 构造算法的具体描述过程如算法 1 所示。

### 算法 1 SWAFP-T 构造算法

Function  $SWAFP-T_A$  CONSTRUCTION PROCESS

INPUT:  $VDS, MinSUP, MinallConf, 目标窗, 批, 序列, 初始排序表 ISO;$

OUTPUT: 构成  $SWAFP-T$  树;

生成  $F$ -list;

生成前缀树为空的  $SWAFP-T_A$ ;

FOR 每个批  $B_i$  do

    删除当前窗中孤立的批

    FOR 每个序列  $E_k$  do

        排列  $E_k$  的项;

批次	时间戳	序列
B <sub>1</sub>	1	f <sub>2</sub> f <sub>3</sub> f <sub>6</sub> f <sub>7</sub> f <sub>8</sub>
	2	f <sub>1</sub> f <sub>2</sub> f <sub>4</sub> f <sub>7</sub>
B <sub>2</sub>	3	f <sub>1</sub> f <sub>2</sub> f <sub>3</sub> f <sub>4</sub> f <sub>7</sub> f <sub>8</sub>
	4	f <sub>1</sub> f <sub>3</sub> f <sub>6</sub>
B <sub>3</sub>	5	f <sub>1</sub> f <sub>2</sub> f <sub>4</sub> f <sub>5</sub>
	6	f <sub>1</sub> f <sub>2</sub> f <sub>3</sub> f <sub>4</sub> f <sub>7</sub>
B <sub>4</sub>	7	f <sub>1</sub> f <sub>3</sub>
	8	f <sub>1</sub> f <sub>2</sub> f <sub>7</sub>
	-	--

图 1 振动数据流结构示意图

Fig.1 Structure diagram of vibration data flow

```

更新F-list中频率支持度;
将 $E_k$ 插入到SWAFP-TA;
END FOR
按频率降序从F-list中计算 $F_{fd}$ ;
FOR SWAFP-TA 的每个分支 do
    使用BSM算法重构压缩;
END FOR
END FOR
END
    
```

如图1所示,数据流被划分为4个批,每个批包含相同数目的序列,窗则由固定数目的非重叠批组成,图1中的每个窗有3个批。图2给出了窗1的SWAFP-T树构建过程。如图2(a)所示,SWAFP-T树初始化为空,并从“null”根节点开始构建。TS=1序列  $\{f_2 f_5 f_6 f_7 f_8\}$  按顺序  $\{\{\} \rightarrow f_2:1 \rightarrow f_5:1 \rightarrow f_6:1 \rightarrow f_7:1 \rightarrow f_8:1\}$  插入到树中,第一分支中 $f_2$ 为根节点之后的初始节点而 $f_8$ 为尾节点, $f_x:1$ 表示频率 $f_x$ 的支持度值为1,如定义1所述。在插入第2条序列前,为维护F-list,将TS=2的频率排列顺序从 $\{f_1 f_2 f_4 f_7\}$ 变为 $\{f_2 f_5 f_6 f_7 f_8 f_1 f_4\}$ ,并将其插入到树中。图2(b)为插入批1后的树结构,图2(c)为插入批2的树结构,图2(d)为插入批3的树结构。重构压缩后的树结构如图2(f)所示。

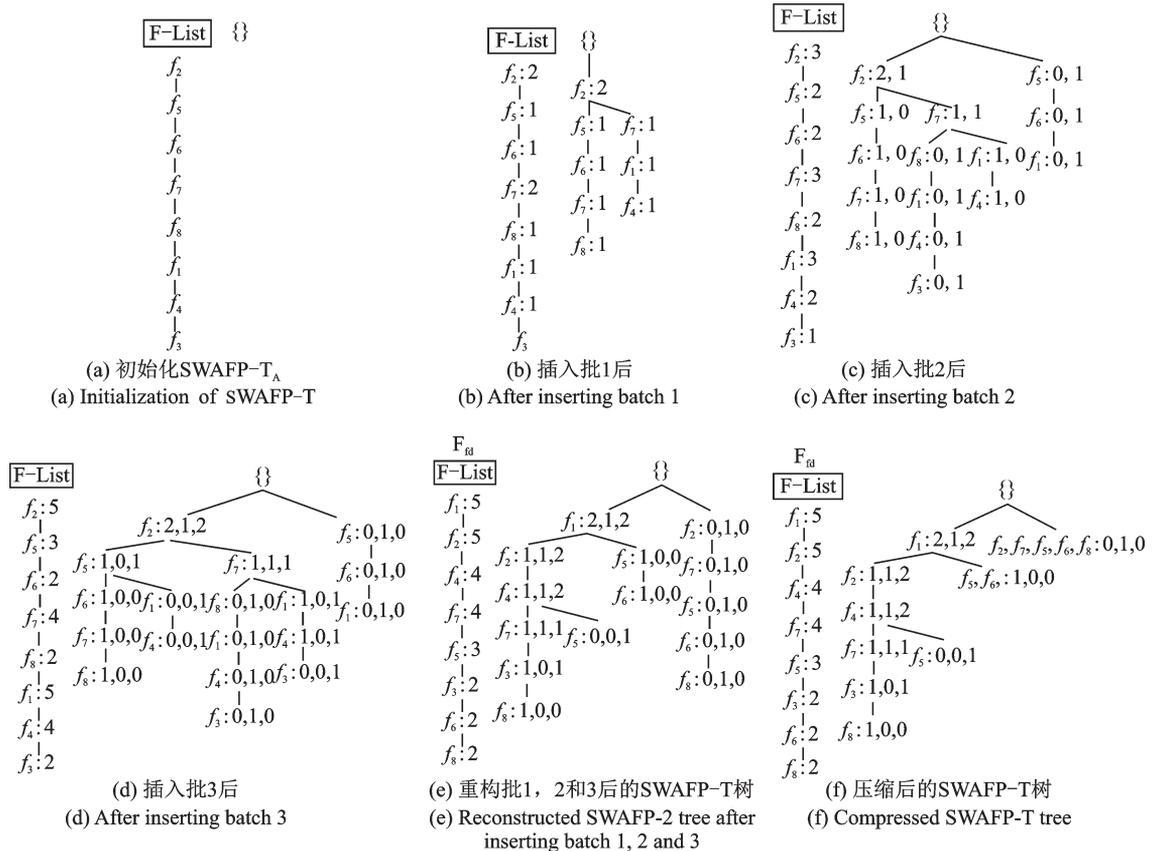


图2 窗1 SWAFP-T树构建过程

Fig.2 Construction process of SWAFP-T in window 1

当振动数据流移动到批4时,由于窗2不再包含批1,因此需要移除批1信息。当有些节点不包含批2和批3的信息时将被从树上去除。图3给出了窗2树构建过程,其中图3(a)删去了批1,图3(b)插入了批4,图3(c)是窗2最终构成的SWAFP-T树。

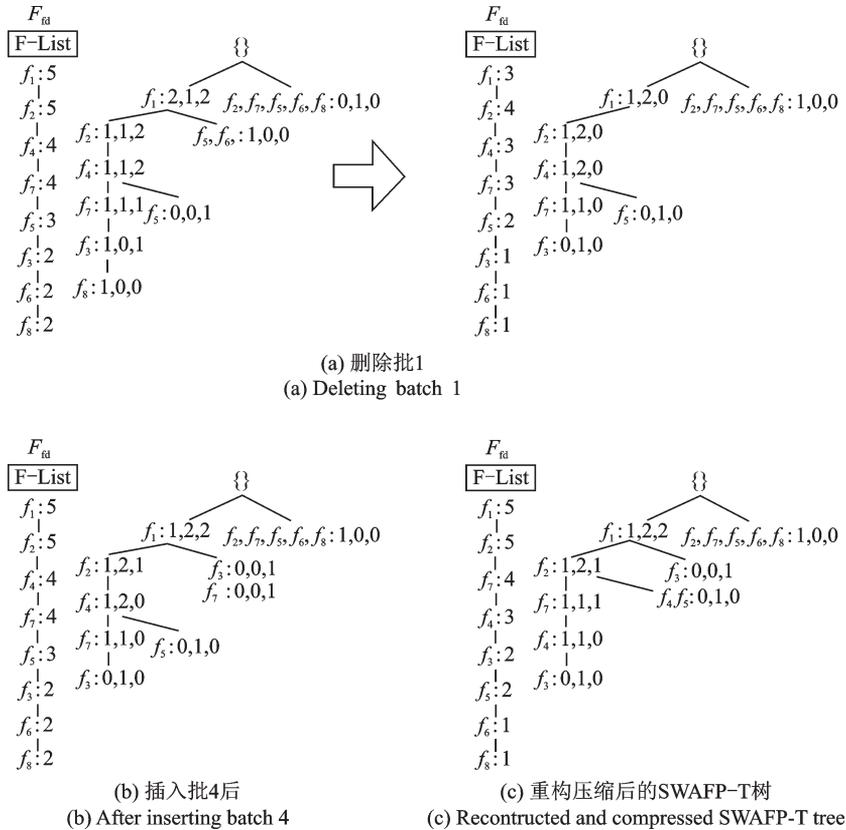


图3 窗2 SWAFP-T树构建过程

Fig.3 Construction process of SWAFP-T in window 2

### 2.3 SWAFP树挖掘过程

从SWAFP-T树上挖掘关联频繁模式的基本操作包括:(1)记录长度为1的频繁频率点;(2)构建每个频率的条件模式基;(3)对每个条件模式基构建条件树。(4)从条件树生成关联模式集,这些过程总结起来得到挖掘算法流程如算法2所示。为了从图1给出的数据流挖掘出关联频繁模式,需要寻找窗1中的所有关联频繁模式集。假定  $MinSUP = 2, MinallConf = 0.55$ 。

使用和文献[20,21]中类似的方法从SWAFP-T树上挖掘所有关联频繁模式,从  $F_{id}$  底部频率开始建立条件模式基(Conditional pattern-bases, CPBs)和条件树(Conditional trees, CTs)。首先从图3(c)建立频率  $f_5$  的条件模式基,这是因为最底部的  $f_6$  和  $f_8$  满足  $MinSUP$  阈值,因此从窗2的SWAFP-T树中的两个分支上提取出频率  $f_5$ 。由这些分支形成的路径分别为  $(f_1 f_2 f_4 f_5 : 1)$  和  $(f_2 f_7 f_5 : 1)$ 。因此,考虑将  $f_5$  作为后缀,其相关的生成条件模式基的两个前缀路径则分别是  $(f_1 f_2 f_4 : 1)$  和  $(f_2 f_7 : 1)$ ,其条件模式基树见图4(a)。如图所示,条件树为空,频率  $f_1, f_4$  和  $f_7$  不在条件树中,这是因为其支持度值为1小于阈值  $MinSUP$ ,并且  $f_1 f_5, f_4 f_5$  和  $f_5 f_7$  的所有信任度值均小于  $MinallConf$ 。另一方面,尽管频率  $f_2$  满足阈值

MinSUP,但是其 $f_2 f_5$ 的信任度值小于MinallConf,所以频率 $f_2$ 也不在条件树中。因此,频率 $f_5$ 没有关联频繁模式。频率 $f_3$ 的条件模式基是 $(f_1 f_2 f_7 f_4: 1)$ 和 $(f_1: 1)$ ,如图4(b)所示。频率 $f_4$ 的条件模式基是 $(f_1 f_2 f_7: 1)$ 和 $(f_1 f_2: 1)$ ,其条件树只包含一个单路径 $(f_1 f_2: 3)$ ,如图4(c)所示。 $f_4$ 的关联频繁模式集有 $(f_1 f_4: 1)$ 、 $(f_2 f_4: 3)$ 和 $(f_1 f_2 f_4: 3)$ 。表1给出了图1数据流在窗2下的关联频繁模式集的挖掘结果。

**算法 2** SWAFP-T 挖掘算法

Function Mining\_process of SWAFP-T

Begin

For  $f_i$  in F-list do

从底部开始寻找长度为1的 $f_i$ ;

For each  $f_j$  do

If 位于 $f_i$ 底部的频点 $f_j$ 支持

$< \text{MinSUP}$

删除 $f_j$ ;

以 $f_i$ 为后缀,将前缀路径作为CPBs;

End

For each  $f_j$  in CTs of  $f_i$  do

If  $f_j$  支持度  $\geq \text{MinSUP}$  &&  $f_j f_i$

的信任度  $> \text{MinallConf}$

保留 $f_j$ ;

End;

所有保留元素 $f_j$ 构成CTs;

End for;

For all CTs of  $f_i$  do

求出所有关联频繁模式集;

End for

End for

End for

End

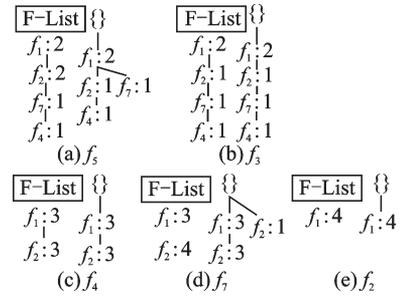


图4 前缀和条件树构建过程

Fig.4 Prefix and conditional tree construction process

表1 窗2下的数据流关联频繁模式集挖掘结果

Tab. 1 Data stream association frequent pattern set mining results in window 2

频率	条件模式基	条件树	关联频繁模式集
$f_3$	$\{(f_1 f_2 f_7 f_4: 1), (f_1: 1)\}$		
$f_5$	$\{(f_1 f_2 f_4: 1), (f_2, f_7: 1)\}$		
$f_4$	$\{(f_1 f_2 f_7: 2), (f_1 f_2: 1)\}$	$\langle f_1 f_2: 3 \rangle$	$f_1 f_4: 3$ $f_2 f_4: 3$ $f_1 f_2 f_4: 3$
$f_7$	$\{(f_1 f_2: 3), (f_2: 1)\}$	$\langle f_1: 3, f_2: 4 \rangle$	$f_1 f_7: 3$ $f_2 f_7: 4$ $f_1 f_2 f_7: 3$
$f_2$	$\{f_1: 4\}$	$\langle f_1: 4 \rangle$	$f_1 f_2: 4$

**3 故障状态监测**

本节在AFPMF框架下利用SWAFP-T树进行故障概率识别,从而实现故障状态监测。首先计算故障特征频率,主要过程如下:检查频率信息表(表2),对比具有可疑振动频率信息的轴承ID,此具体频率信息由式(1—3)计算。若表中可疑振动频率出现在关联频繁模式集SWAFP-T树中,则认为该振动频率对应的器件有可能出现故障。以轴承1为例,若其轴速为350圈/min,则内圈故障

表2 轴承故障频率

Tab. 2 Bearing failure frequency

轴承ID	$f_{it}$	$f_{oi}$	$f_{bt}$	Hz
1	10.65	8.52	3.68	
2	11.15	9.65	4.26	

频率为  $10.65 \times 350 = 3\,727.5$  圈/min = 62.125 Hz。

一旦 SWAFP-T 树中出现以上频率或者相应的谐波频率成分,则认为此轴承内圈有一定概率存在故障,进而启动相应的修复进程替换该潜在故障轴承。

## 4 实验结果分析

为识别故障频率,需要从故障轴承数据中挖掘所有的关联频繁模式集。实验分析了3种不同故障如轴承内圈、外圈和滚珠故障条件下的振动数据,数据收集使用16通道的数字式磁带录音机,采样速率为12 000次/s。在AFPMF数据预处理过程中,将获得的每种故障振动数据通过Matlab软件FFT函数转换为频域数据。AFPMF中数据挖掘算法和故障状态监测均使用Matlab 2012b实现,软件运行硬件环境的CPU是2.8 GHz,内存为8 GB。

### 4.1 算法运行效率

首先,评估不同时间窗大小和不同批大小条件下AFPMF算法的运行效率。先将内圈数据集划分为12批,每批包含100条序列,窗口大小设为  $W=3B$ 。在滑动窗口环境下,窗1包含了最开始的3个批次,分别为  $B_1, B_2$  和  $B_3$ 。同理,窗2包含的批分别是  $B_2, B_3$  和  $B_4$ 。再次重复实验时,数据划分为6批,每批包含200条序列,同样窗口大小为  $W=3B$ 。图5(a)给出了不同阈值 MinallConf 时算法应用于内圈数据的执行时间对比,其中设置 MinSUP 为 30%。图5(b)给出了批大小分别为150和300条序列并且窗口  $W=3B$  的外圈数据挖掘算法结果,实验中 MinSUP 固定设置为 40%。由图5结果可知,算法运行时间在所有数据集下均会随着 MinallConf 的增加而下降。其次,图6给出了算法执行时间在 MinallConf 固定、MinSUP 不同时的实验结果对比。从图6中可以看出,所有类型数据集下算法的运行时间均会随着 MinSUP 的增加而下降,这是因为候选模式集数目减小所致。

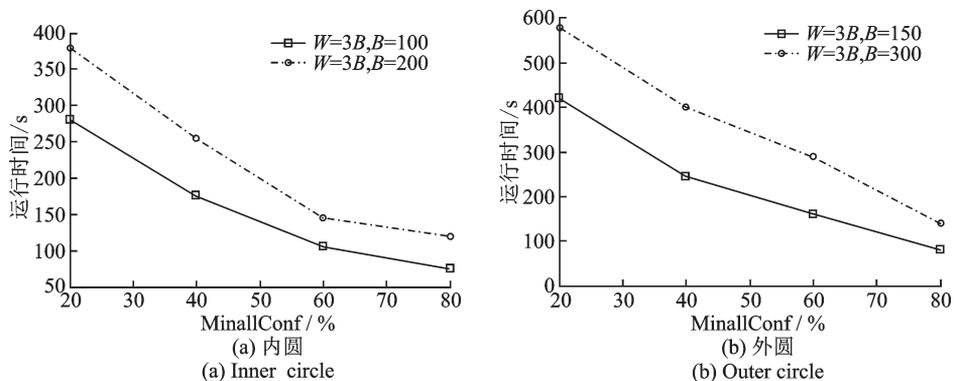


图5 不同 MinallConf 时算法挖掘时间对比结果

Fig.5 Experimental results of mining time at different Minallcof

### 4.2 故障诊断识别性能

为识别故障频率,实验采用驱动端轴承,数据参数如表3所示,故障频率计算结果如表4所示。根据第3节中故障频率的计算公式,若表中可疑振动频率出现在关联频繁模式集 SWAFP-T 树中,则认为该振动频率对应的器件有可能出现故障。通过计算可知内圈、外圈和滚珠故障频率分别为 158, 210 和 138 Hz。首先,检查在关联模式集中是否会出现故障频率和谐波成分。如果出现某类型故障频率,则说明该数据集对应的轴承部件有可能出现故障。实验结果显示,关联模式集中出现 316 Hz 频率,由于这

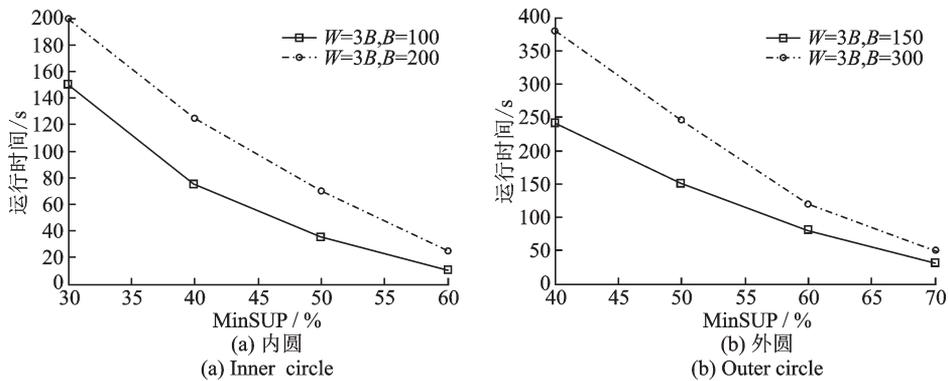


图6 不同MinSUP时算法挖掘时间对比结果

Fig.6 Experimental results of mining time at different MinSUP

是158 Hz的谐波频率,因此轴承内圈可能发生故障。本方案的优点在于只要有任何故障频率存在,不管阈值如何设置,SWAFP-T树均能够识别出故障频率。

表3 轴承主要特征参数

内径	外径	厚度	滚珠直径	节圆直径
0.948 2	2.122 3	0.612 3	0.321 4	1.542 1

表4 故障频率

内圈	外圈	滚珠
158	210	138

### 4.3 故障诊断精度

为评估不同高斯白噪声干扰(0 dB到-10 dB)条件下算法对内圈、外圈和滚珠数据的故障诊断精度,定义故障诊断精度为加噪声后检测到的故障频率数目和加噪声前检测数目之比。

以内圈数据为例,未加噪声时,主要参数有 $W=3B, B=100, MinallConf=55\%, MinSUP=45\%$ 。实验得到了60个与316 Hz频率相关联的可疑故障频率,其中316 Hz是4.2节158 Hz的故障谐波频率。加过噪声之后,在0和-2 dB时实验结果相同,而对于-4, -6, -8, -10 dB时分别得到59, 57, 56和55个关联频率数目。内圈、外圈和滚珠的实验结果如表5所示,其中外圈和滚珠的MinSUP和MinallConf分别为55%和65%, 60%和70%。

本文将文献[6,8,9,11]中提到的4种算法作为对照算法,通过适当改进应用于振动数据流,结果如表6。据表6可知,AFPMF识别故障的精度显著优于现有技术。

表5 SWAFP-T算法精度

Tab. 5 Accuracy of SWAFP-T algorithm

实验序号	SNR /dB	精确度/%		
		内圈	外圈	滚珠
1	0	100	100	100
2	-2	100	100	100
3	-4	98	97	100
4	-6	94	95	97
5	-8	92	94	95
6	-10	91	90	93

表6 SWAFP和已有算法的故障识别精度实验结果对比

Tab. 6 Experiments on fault recognition accuracy of SWAFP and existing algorithms

方法	文献 [9]	文献 [8]	文献 [6]	文献 [11]	AFPMF
精度/%	76	38	85	90	93

## 5 结束语

本文提出了一种应用于轴承故障数据场景的数据挖掘框架AFPMF。在

AFPMF框架中,首先使用FFT将振动数据流从时域转换为频域数据,再使用基于滑动窗口的挖掘算法SWAFP寻找到关于故障频率信息的关联频繁模式集,使用了具有高压缩、高效率可重构的SWAFP-T树用于关联频繁模式集的挖掘。最后,将关联频繁模式集直接应用于故障识别,实现故障状态监测。实验结果表明,本文所提方案有较快的算法执行效率,并且拥有比现有算法更优越的故障识别精度。

#### 参考文献:

- [1] 李巍华,单外平,曾雪琼.基于深度信念网络的轴承故障分类识别[J].振动工程学报,2016(2):12-19.  
Li Weihua, Shan Waiping, Zeng Xueqiong. Bearing fault identification based on deep belief network[J]. Journal of Vibration Engineering, 2016(2): 12-19.
- [2] Yaqub MF, Gondal I, Kamruzzaman J. Multi-step support vector regression and optimally parameterized wavelet packet transform for machine residual life prediction [J]. Journal of Vibration & Control, 2012, 19(7): 963-974.
- [3] Bellini A, Filippetti F, Tassoni C, et al. Advances in diagnostic techniques for induction machines[J]. IEEE Transactions on Industrial Electronics, 2008, 55(12): 4109-4126.
- [4] 杜振宁,向春枝.基于小波包分解和PCA的轴承故障诊断[J].控制工程,2016,23(6):812-815.  
Du Zhenning, Xiang Chunzhi. A wavelet packet decomposition and principal component analysis approach for feature extraction in bearing failure vibration signal[J]. Control Engineering of China, 2016, 23(6): 812-815.
- [5] Yaqub M F, Gondal I, Kamruzzaman J. Inchoate fault detection framework: Adaptive selection of wavelet nodes and cumulant orders [J]. IEEE Transactions on Instrumentation & Measurement, 2012, 61(3): 685-695.
- [6] Li F, Meng G, Ye L, et al. Wavelet transform-based higher order statistics for fault diagnosis in rolling element bearings[J]. Journal of Vibration & Control, 2008, 14(11): 1691-1709.
- [7] 郝如江,卢文秀,褚福磊.滚动轴承故障信号的数学形态学提取方法[J].中国电机工程学报,2008,28(26):65-70.  
Hao Rujiang, Lu Wenxiu, Chu Fulei. Mathematical morphology extracting method on roller bearing fault signals[J]. Proceedings of the CSEE, 2008, 28(26): 65-70.
- [8] Malhi A, Gao R X. PCA-based feature selection scheme for machine defect classification[J]. IEEE Transactions Instrumentation & Measurement, 2008, 53(6): 1517-1525.
- [9] 张应红,李聪,景晖,等.基于人工神经网络技术的矿用皮带滚动轴承故障诊断[J].机床与液压,2014,42(3):180-183.  
Zhang Yinghong, Li Cong, Jing Hui, et al. Fault diagnosis of roller element bearing of mine belt conveyor based on artificial neural network technique[J]. Machine Tool & Hydraulics, 2014, 42(3): 180-183.
- [10] 张洪,许鸿,唐侃,等.基于PLS的滚动轴承装配质量预测[J].控制工程,2016,23(3):332-335.  
Zhang Hong, Xu Hong, Tang Kan, et al. Rolling bearing assembly quality forecasting based on partial least squares[J]. Control Engineering of China, 2016, 23(3): 332-335.
- [11] 李巍,徐雷.基于数据挖掘技术的医疗设备故障库的设计[J].中国医疗设备,2016,31(5):269-272.  
Li Wei, Xu Lei. Design of medical equipment failures library based on data mining technology[J]. China Medical Devices, 2016, 31(5): 269-272.
- [12] Rakesh A, Ramakrishnan S. Fast algorithm for mining association rules in large databases [C]//Proceedings of the 20th International Conference on Very Large Data Bases.[S.l.]: [s.n.], 1993: 207-216.
- [13] Han J, Pei J, Yin Y, et al. Mining frequent patterns without candidate generation: A frequent-pattern tree approach [J]. Data Mining and Knowledge Discovery, 2004, 8: 53-87.
- [14] Giannella C, Han J, Pei J. Mining frequent patterns in data streams at multiple time granularities[C]// Conference of Next Generation Data Mining. Cambridge: MIT Press, 2003: 191-212.
- [15] Arasu A, Manku G S. Approximate counts and quantiles over sliding windows[C]//Proceedings of 23rd ACM

SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems. Paris, France: ACM, 2004: 286-296.

- [16] Yun C, Haixun W, Philip S Y. Catch the moment maintaining closed frequent itemsets over a data stream sliding window [J]. Knowledge and Information Systems, 2006, 10(3): 265-294.
- [17] Amar M, Gondal I, Willson C. Wibration spectrum imaging: A novel bearing fault classification approach [J]. IEEE Transactions on Industrial Electronics, 2014, 62(99): 494-502.
- [18] Tanbeer Sk, Ahmed CF, Jeong BS. Efficient single pass frequent pattern mining using a prefix-tree[J]. Information Science, 2009, 179: 559-583.
- [19] Leung C K S, Khan Q I, Li Z, et al. Can-tree: A canonical-order tree for incremental frequent pattern mining [J]. Knowledge Information System, 2007, 11: 287-311.
- [20] Rashid M M, Gondal I, Kamruzzaman J. Mining associated patterns from wireless sensor networks [J]. IEEE Transactions on Computing, 2015, 64(7): 1998-2011.
- [21] Rashid M M, Gondal I, Kamruzzaman J. Mining associated sensor patterns for data stream of wireless sensor networks[C]// Proceedings of 8th ACM Workshop on Performance Monitoring and Measurement of Heterogeneous Wireless and Wired Network. [S.l.]: ACM, 2013: 91-98.

作者简介:



张艳梅(1975-),女,硕士,讲师,研究方向:计算机网络与数据挖掘, E-mail: 29064248@qq.com。

陆伟(1978-),男,博士研究生,研究方向:计算机网络和云计算。

杨余旺(1966-),男,博士,教授,博士生导师,研究方向:物联网和网络编码算法及应用技术。

(编辑:王静)