

# 数据驱动的动车组滚动轴承故障预测

李 莉<sup>1</sup>, 贾志凯<sup>1</sup>, 张 瑜<sup>2</sup>, 李时法<sup>3</sup>

(1. 中国铁道科学研究院, 北京 100080; 2. 北京经纬信息技术公司, 北京 100080;  
3. 新乡市起重设备厂有限责任公司, 河南 新乡 453003)

**摘要:** 为了有效提高动车组滚动轴承故障的发现率, 减少故障监控系统的误报现象, 基于 Apache Hadoop 大数据平台对经典 Apriori 算法进行改进, 并将其应用于动车组滚动轴承故障的预测研究工作中。首先, 针对经典 Apriori 算法的不足, 在 MapReduce 框架下提出以业务经验为约束的改进的 Apriori 算法。其次, 基于文中提出的改进的 Apriori 算法对某铁路局的动车组状态、故障预警、维修历史等信息进行深度数据挖掘, 并通过得出的关联规则进行动车组滚动轴承故障的预测。实验结果表明, 文中提出的算法准确率达 72%, 减少了 80% 以上的误报报警信息, 在实验环境中运算效率较传统的 Apriori 算法提高了 50%。

**关键词:** 智能交通; 故障预测; Apriori 算法; 数据挖掘; 大数据

中图分类号: U279 文献标志码: A 文章编号: 1672-3767(2017)04-0016-08

DOI: 10.16452/j.cnki.sdkjzk.2017.04.003

## Antifriction Bearing Failure Prediction of EMU Based on Data Driven Approach

LI Li<sup>1</sup>, JIA Zhikai<sup>1</sup>, ZHANG Yu<sup>2</sup>, LI Shifa<sup>3</sup>

(1. China Academy of Railway Science, Beijing 100080, China;  
2. Beijing Jingwei Information Technologies Company, Beijing 100080, China;  
3. Xinxiang Hoisting Equipment Factory, Xinxiang, Henan 453003, China)

**Abstract:** In order to effectively increase the antifriction bearing failure discovery rate in EMU and reduce the failure misinformation, the classical Apriori algorithm was improved based on the Apache Hadoop big data platform, and the improved algorithm was applied into the research work of EMU antifriction bearing failure prediction. To overcome the limitations of classical Apriori algorithm, an improved Apriori algorithm under the constraints of professional experience was proposed in the MapReduce framework. Then the depth data such as the status, failure warning and maintenance history of a certain railway bureau's EMU were mined based on the improved Apriori algorithm and the prediction of EMU antifriction bearing failures was obtained based on some association rules. The results show that with an accuracy rate of 72%, the failure misinformation of the proposed algorithm is decreased by 80%, and compared with the classical Apriori algorithm, the computing efficiency of the proposed algorithm is increased by 50%.

**Key words:** intelligent transportation; failure prediction; Apriori algorithm; data mining; big data

中国高速动车组运行速度高, 运行里程大, 连续运行时间长, 走行部载荷较大, 故障率较高。据统计, 自 2010 年下半年至 2016 年底, 在全路动车组范围内转向架运行故障仅次于电力电子部件和车内设施。而滚

收稿日期: 2017-04-05

基金项目: 中国铁道科学院院基金(2016TJ102)

作者简介: 李 莉(1979—), 女, 河南新乡人, 博士研究生, 主要研究智能管理信息系统, 本文通信作者。

E-mail: lili2005@mail.lzjtu.cn.

贾志凯(1975—), 男, 河北新河人, 副研究员, 主要研究铁路计算机应用技术。

张 瑜(1981—), 男, 辽宁铁岭人, 工程师, 主要研究高速铁路信息化。

李时法(1965—), 男, 河南新乡人, 高级工程师, 主要研究机械设备故障分析技术。

动轴承作为转向架的重要部件,故障频发。由于目前动车组轴承故障缺少地面监测设备,车载网络及监控系统<sup>[1,2]</sup>存在误报漏报现象,统计结果显示,平均一组重连车日报警量多达 300 多次,误报比例高,给现场工作人员造成很大困扰,给动车组运行安全带来隐患,严重影响高速铁路正常运行秩序,甚至会导致安全事故的发生。因此利用单点实时监测数据及报警信息综合分析滚动轴承未来发生故障的可能性,提高故障发现率,减少监控系统误报,具有重要意义。

目前用于复杂装备的故障趋势分析及预测的主要方法有基于故障机理、基于故障征兆、数据驱动的,采用神经网络、模糊逻辑等方法,取得了一些较为突出的成果<sup>[3-8]</sup>。文献[3-5]通过实际试验或基于故障机理的模型进行故障预测,预测方法对专业水平要求高,建立模型时不可避免的忽略次要因素,无法全面考虑故障的影响因素,影响预测精度。文献[6]对不同的数据驱动故障预测方法进行评估,得出即使不同方法损伤估计不同,却都可以对剩余寿命进行预测的结论。文献[7]采用不同模型获得数据进行预测研究,对于不同研究对象和不同数据特征均有不同的较为适合的方法。文献[8]采用了神经网络和模糊逻辑相结合的方法建模,利用在线数据进行测试,未考虑利用历史数据进行较长一段时间内的故障预测。动车组滚动轴承服役状态复杂,故障的发生受运行速度、运行时间、地形、气候等多种因素影响,采用传统的基于故障机理等故障预测方法难以建立精确故障预测模型。目前尚缺少行之有效的能够实际应用的故障预测模型。纵观动车组故障历史,故障发生后,追溯系统内故障发生前的动车组运行数据,总能发现蛛丝马迹。经典的 Apriori 算法能够准确的发现不同特征向量间的关联规则,然而其运算效率偏低、占用计算资源偏多。

本文提出一种改进的 Apriori 算法,并利用其分析滚动轴承故障征兆的关键参数,即采用基于故障征兆与基于数据相结合的方法进行动车组滚动轴承故障的预测。从数据角度出发,综合分析动车组部件履历、状态监测、故障预警、故障处理、维修等历史数据,揭示动车组故障发生规律,进行故障预警和早期故障预测。

## 1 经典 Apriori 算法及其存在的不足

### 1.1 经典 Apriori 算法

Apriori 算法<sup>[9-10]</sup>是一种最早的最有影响力的挖掘布尔关联规则的频繁项集的数据挖掘算法。基本思想是循环扫描全部事务数据,逐层搜索。利用迭代的方法,项目集元素个数不断增多,逐步发现所有频繁项集。首先第一次扫描,产生项目集个数为 1 的候选项集  $C_1$ ,从中产生不小于最小支持度的事务记录形成频繁项目集  $L_1$ ;然后是 2-频繁项目集  $L_2$ ,直到不能再扩展频繁项目集元素数目算法停止。在第  $k$  次循环中,利用  $L_{k-1}$ 通过两两连接,产生  $k$ -候选项目集的集合  $C_k$ , $C_k$ 中各事务记录与  $L_{k-1}$ 中有  $k-1$ 项相同,只有最后一项不同。在此过程中将含有非频繁项的项目集剪除。然后通过扫描数据库生成支持度并测试产生  $k$ -频繁项目集  $L_k$ ,从而获得所需频繁模式。最后利用置信度产生关联规则。作为经典数据挖掘算法,Apriori 算法发展比较成熟,在商业和网络安全等领域中应用广泛<sup>[11-12]</sup>。

关联规则形如  $X \rightarrow Y, X \subseteq D, Y \subseteq D, X \cap Y = \Phi$ ,必须同时满足支持度和置信度要求。

定义 1 支持度(Support)是数据集  $D$ 中包含  $X \cup Y$ 的事务所占的百分比,即

$$Support(X \rightarrow Y) = P(X \cup Y) = |T: X, Y \subseteq T, T \subseteq D| / |D| \quad (1)$$

定义 2 置信度(Confidence)是指在数据集  $D$ 中包含  $X$ 的事务中,包含  $Y$ 的百分比,即

$$Confidence(X \rightarrow Y) = P(Y | X) = Support(X \rightarrow Y) / Support(X) \quad (2)$$

事务数据库  $D$ 中所有不小于最小支持度的非空项目集,称为频繁模式。若项目集是非频繁项目集,则其所有超集都是非频繁项目集,应剪除。

### 1.2 经典 Apriori 算法中存在的不足

由于 Apriori 算法需要多次扫描数据库,并产生大量的候选项集,当利用传统数据处理平台进行大数据级的数据挖掘时,大数据量存储、内存需求、处理速度等问题将变得尤为突出。截至目前,全路已建成 56 个动车所,7 个动车基地。据统计,每个动车运用所累积数据已达 20 G 左右,全路数据量达到 TB 级,其中包括大量滚动轴承故障相关数据。随着中国高速铁路快速发展,动车组运行安全监控广度和深度将不断加强,动车组滚动轴承的运行状态监测数据和有关故障记录及处理数据也将越来越多。采用经典 Apriori 算法需要多次扫描数据

库,经测算,对于项集最大长度为 7,数据库规模为 30 万条数据,运算时长超过 4min。动车组停车时间一般为 2 min,因此当预测发现紧急情况时,没有足够的时间进行处理,无法满足在线的故障预测应用要求。除此之外随着项目集元素最大个数的增大,候选项集将急剧增多,如果 1-项频繁集的数量为  $10^4$ ,则 2-项候选集将达到  $10^7$ ,并且每个候选项集都需要验证,系统资源开销大,同时也会大大降低系统运算效率。然而现有的经典 Apriori 算法改进成果<sup>[13-15]</sup>无法满足滚动轴承大数据量分析和在线快速预测要求。

## 2 改进 Apriori 算法研究

针对 2.2 中提出的问题,为满足快速处理 TB 级动车组滚动轴承数据的目的,本文对经典 Apriori 算法从以下两个方面改进(图 1):①基于 Hadoop 对经典算法进行改进,使之能够适应 Apache hadoop 的 MapReduce 框架<sup>[16]</sup>,实现并行计算,从而减少运算时间。②在利用支持度剪枝的同时,将业务应用经验作为约

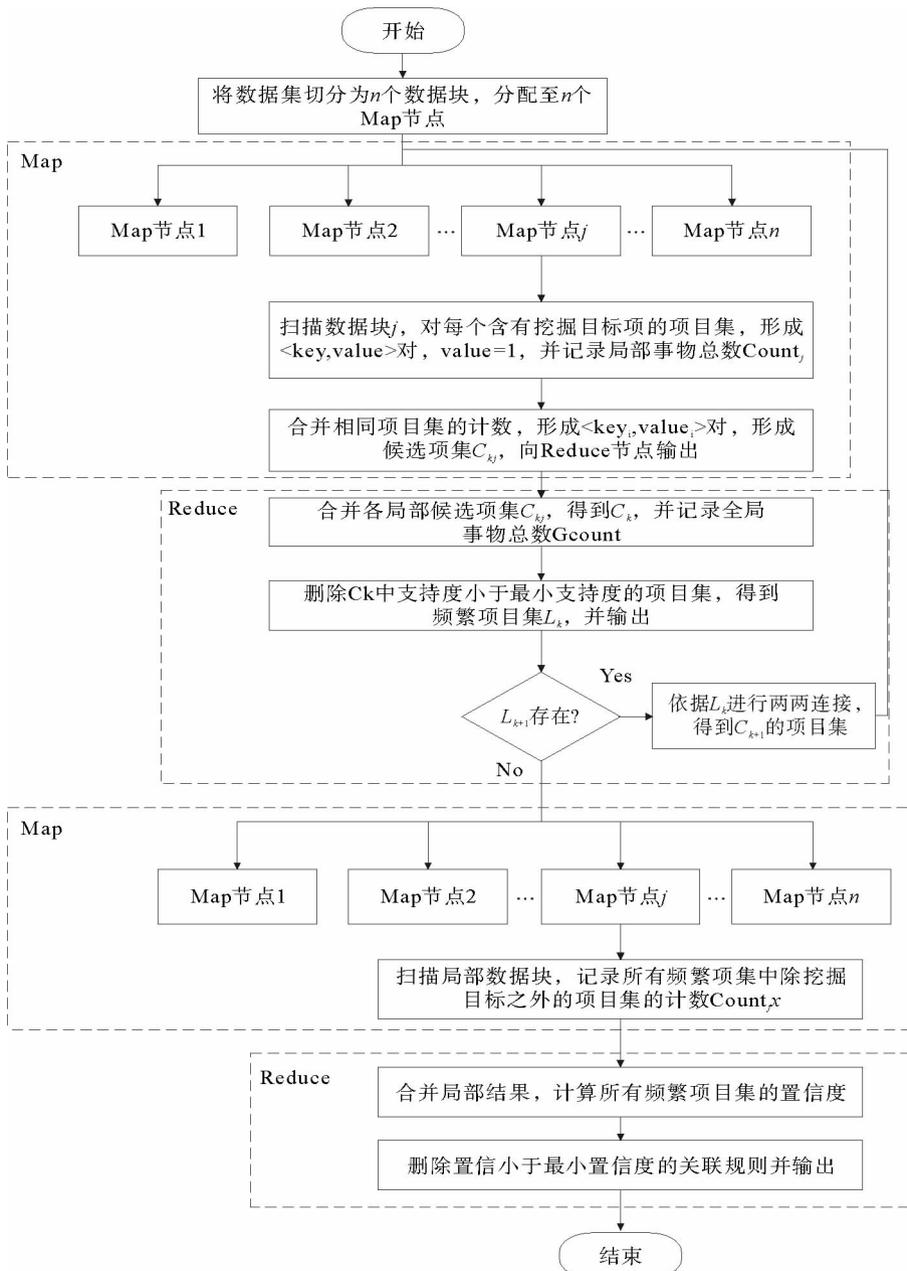


图 1 改进的 Apriori 算法流程图

Fig. 1 Improved Apriori algorithm flow

束条件减少候选项集的产生,即将预测目标作为约束条件,可大幅减少候选项集的产生。

改进的 Apriori 算法将从以下五个步骤展开(如图 1):

1) 将数据集切分为  $n$  个数据块,并分发至不同 Map 节点。

2) 每个 Map 节点处理本地数据,每条事务记录生成  $\langle \text{key}, \text{value} \rangle$ , Key 为项目集,value 为 1。同时记录局部所有事务记录的数目。

3) 每个节点对 Map 结果进行合并,形成  $\langle \text{key}_i, \text{value}_i \rangle$  输出;其中,  $\text{key}_i$  为每个项目集,  $\text{value}_i$  为对应项目集在本地数据块中出现次数。

4) Reduce 节点接收各 Map 节点的输出结果并进行合并,得出每个项目集的在全部数据中的支持计数  $\text{value}_i$  和全局事务记录总数,依次形成  $k$ -项目集  $C_k$  所有满足支持度要求的频繁项目集  $L_k$ 。

5) 计算各频繁项目集的置信度,满足置信度要求的形成关联规则。

### 3 基于改进的 Apriori 算法进行动车组滚动轴承故障预测应用

#### 3.1 改进的 Apriori 算法应用

以某铁路局的动车组运用数据为例。数据源为铁路局辖下的动车基地和运用所的动车组管理信息系统中的相关数据。利用改进后的 Apriori 算法对部件履历、状态监测信息、故障历史信息、维修历史数据、检修故障及其它相关部件故障与滚动轴承近期内发生故障之间的规律进行数据挖掘。同时将动车组运行参数、状态信息和故障预警信息及故障历史信息作为分析项目,假设将动车组某一状态点之后的两个月内通过检修确认的故障(检修故障)作为分析结果项目,即检修故障发生是形成频繁项目的约束条件,此类约束可以降低候选项集的产生,提高挖掘效率。

##### 1) 特征向量提取

通过分析,与滚动轴承紧密关联的轮对若出现裂纹或多边形等问题,会引起动车组运行过程中振动加剧,导致滚动轴承随之振动异常,磨损加剧,继而出现故障概率提高;同样,转向架上悬挂的牵引电机故障也会引起类似后果;动车组的高速运行会引起滚动轴承温度升高,尤其是温度持续升高时磨损会加剧;随着滚动轴承走行公里增加,其健康状态退化,容易发生故障。综上所述,动车组滚动轴承故障的发生同与之紧密关联的转向架其他关键部件的状态、滚动轴承故障历史和动车组运行状态有关。提取的特征向量包括:动车组走行公里、故障发生时动车组运行速度、轴温报警信息、车轮近期历史故障信息、齿轮箱近期历史故障信息、滚动轴承温度持续上升、滚动轴承自身未来近期故障信息。

##### 2) 数据预处理

利用最近七年数据进行数据挖掘。各业务系统的原始数据存在命名规范不统一,故障描述不统一,存在大量非结构化数据等问题。抽取到原始数据后,需要根据时间和动车组编号、轴号等信息进行配准,去重,统一命名,结构化故障描述、故障原因透析等大量工作。数据预处理后得到近三十五万条记录,数据示例如表 1 所示。

其中,

A1:滚动轴承的累计走行公里达到或超过 8 万 km;

A2:事务发生时刻的动车组运行速度达到或超过 120 km/h;

A3:包括以下三种情况发生:3 级轴温报警频率达到或超过 3 次/h,2 级轴温报警频率超过达到或超过 2 次/h,1 级轴温报警;

A4:事务发生时间前 3 个月内车轮发生故障;

A5:事务发生时间前 3 个月内齿轮箱发生故障;

表 1 预处理后数据格式

Tab. 1 Data Format after Preprocessing

序号	A1	A2	A3	A4	A5	A6	A7
1	1	1	0	1	0	0	1
2	1	0	1	0	0	1	0
...	...	...	...	...	...	...	...

0 表示未发生,1 表示发生。

- A6: 轴承箱温度平均值连续二小时持续上升;
- A7: 事务发生后 2 个月内检修确认发生故障。

3) 关联规则发现

搭建实验环境部署 hadoop 运行环境, 共有 5 个 Map 节点, 将数据集平均分为 5 个数据块分发至不同的 Map 节点。每个节点扫描处理本地数据, 然后合并提交至 Reduce 节点进行汇总分析, 这里只关心包含 {A7} 的项目集。由于样本基数大, 而动车组故障样本较少, 故动车组设置最小支持度为 0.5%, 最小置信度 30%, 分析过程如表 2 至表 11 所示。其中表 2、4、6、8、10 为候选项目集  $C_1$  至  $C_5$  及其支持计数, 项目集个数为 1 至 5 (除约束条件 A7 外)。表 3、5、7、9、11 为频繁项集  $L_1$  至  $L_5$ , 为候选项集中支持计数满足最小支持度要求的项目集。表 12 为关联规则, 为频繁项集中置信度满足最小置信度要求的项目集。

表 2 1-项目集  $C_1$

Tab. 2 1-item set  $C_1$

项目集	支持计数
{A1A7}	24 369
{A2A7}	18 012
{A3A7}	10 258
{A4A7}	4 562
{A5A7}	2 245
{A6A7}	13 641

表 3 频繁项目集  $L_1$

Tab. 3 Frequent item set  $L_1$

项目集	支持计数
{A1A7}	24 369
{A2A7}	18 012
{A3A7}	10 258
{A4A7}	4 562
{A5A7}	2 245
{A6A7}	13 641

表 4 2-项目集  $C_2$

Tab. 4 2-item set  $C_2$

项目集	支持计数
{A1A2A7}	14 264
{A1A3A7}	9 087
{A1A4A7}	2 435
{A1A5A7}	1 789
{A1A6A7}	10 697
{A2A3A7}	7 446
{A2A4A7}	2 313
{A2A5A7}	1 789
{A2A6A7}	10 243
{A3A4A7}	2 156
{A3A5A7}	1 369
{A3A6A7}	5 674
{A4A5A7}	231
{A4A6A7}	3 127
{A5A6A7}	1 790

表 5 频繁项目集  $L_2$

Tab. 5 Frequent item set  $L_2$

项目集	支持计数
{A1A2A7}	14 264
{A1A3A7}	9 087
{A1A4A7}	2 435
{A1A5A7}	1 789
{A1A6A7}	10 697
{A2A3A7}	7 446
{A2A4A7}	2 313
{A2A5A7}	1 789
{A2A6A7}	10 243
{A3A4A7}	2 156
{A3A6A7}	5 674
{A4A6A7}	3 127
{A5A6A7}	1 790

表 6 3-项目集  $C_3$   
Tab. 6 3-item set  $C_3$

项目集	支持计数
{A1A2A3A7}	11 423
{A1A2A4A7}	2 765
{A1A2A5A7}	1 678
{A1A2A6A7}	11 457
{A1A3A4A7}	653
{A1A3A5A7}	346
{A1A3A6A7}	6 897
{A1A4A5A7}	96
{A1A4A6A7}	6 975
{A1A5A6A7}	875
{A2A3A4A7}	2 098
{A2A3A5A7}	1 124
{A2A3A6A7}	4 353
{A2A4A5A7}	2 347
{A2A4A6A7}	982
{A2A5A6A7}	941
{A3A4A5A7}	838
{A3A4A6A7}	1 135

表 7 频繁项目集  $L_3$   
Tab. 7 Frequent item set  $L_3$

项目集	支持计数
{A1A2A3A7}	11 423
{A1A2A4A7}	2 765
{A1A2A5A7}	1 678
{A1A2A6A7}	11 457
<b>{A1A3A6A7}</b>	<b>6 897</b>
{A1A4A6A7}	6 975
{A2A3A4A7}	2 098
<b>{A2A3A6A7}</b>	<b>4 353</b>
{A2A4A5A7}	2 347

表 8 4-项目集  $C_4$   
Tab. 8 4-item set  $C_4$

项目集	支持计数
{A1A2A3A4A7}	6 547
{A1A2A3A5A7}	1 732
{A1A2A3A6A7}	7 656
{A1A2A4A5A7}	1 078
{A1A2A4A6A7}	1 674
{A1A2A5A6A7}	1 432
{A2A3A4A5A7}	547
{A2A3A4A6A7}	1 847
{A2A4A5A6A7}	1 858

表 9 频繁项目集  $L_4$   
Tab. 9 Frequent item set  $L_4$

项目集	支持计数
<b>{A1A2A3A4A7}</b>	<b>6 547</b>
{A1A2A3A5A7}	1 732
<b>{A1A2A3A6A7}</b>	<b>7 656</b>
{A1A2A4A6A7}	1 674
<b>{A2A3A4A6A7}</b>	<b>1 847</b>
{A2A4A5A6A7}	1 858

表 10 5-项目集  $C_5$   
Tab. 10 5-item set  $C_5$

项目集	支持计数
{A1A2A3A4A5A7}	635
{A1A2A3A4A6A7}	2 567
{A1A2A3A5A6A7}	1042

得到关联规则 6 条(表 7、表 9、表 11 中黑体标示部分),6 条规则及相应的置信度见表 12。

表 12 中,规则 1 表明在系统故障报警时,轴承箱温度呈持续升高趋势,在滚动轴承的走行公里达到或超过检修周期的 80%,同时车轮在近期发生过故障,并且动车组速度达到 120 km/h 以上时,发生滚动轴承故障的概率高达 71.70%,应当依据具体故障情况立即确认、处理;规则 2、3、4 表明当系统有故障报警时滚动轴承的走行公里达到或超过检修周期的 80%或速度达到 120 km/h 以上时,若轴承箱温度呈持续上升趋势,则滚动轴承发生故障的概率也高达 50%左右;规则 5、6 表明在系统故障报警时,滚动轴承的走行公里达到

或超过检修周期的 80% 或速度达到 120 km/h 以上时,若轴承箱温度呈持续上升趋势则滚动轴承发生故障的概率达到 30% 以上。预测结果说明走行公里未达到 8 万 km 或在低速状态下,发生滚动轴承故障情况相对较少;滚动轴承故障与系统报警关联性最大,其次是走行公里和运行速度,再次是近期内滚动轴承本身故障与车轮故障情况,而与齿轮箱故障间没有强关联关系。

在进行实时处理数据时,表 12 中的单点报警信息为确认报警信息,在系统中可重点提示,其余 80% 以上报警信息均为误报,可以不进行处理。因此可大大减少系统误报率,减少工作人员工作量,提高系统应用效果。

### 3.2 动车组滚动轴承故障预测的实验结果分析

利用 2015 年动车组实际应用数据验证预测规则,准确率达 72%,具有一定的实际应用价值。相对于经典 Aprior 算法,应用在 MapReduce 框架下的改进后 Aprior 算法,在处理小数据量数据时,效率反而较低,但当数据量超过 10 万条后,运算效率优势逐渐明显,当数据量达到 30 万条时,运算效率提高了 50% 左右;明确将未来发生故障作为关联规则约束条件,减少项目集元素个数和候选项集个数,从而进一步降低算法时间复杂度,进一步提高数据挖掘效率,相同条件下运行时间减少 15% 左右。不同数据量运行在不同的数据处理平台的运算时间如图 2 所示。

## 4 结论

在经典 Apriori 算法基础上进行改进,使之适应 Apache hadoop 的 MapReduce 框架,适应大数据量的动车组故障信息数据挖掘,得出较为合理的关联规则,具有较高的故障预测准确度,运算速度能够满足在线故障检测需求,对动车组故障预测具有一定参考意义。随着动车组运行安全监控体系的不断完善,红外轴温、声学早期预警、振动监测、轨道监测等设备的应用,利用本方法进行故障预测,预测准确度可以进一步得到提高。将大数据分析方法与基于故障机理等故障预测方法

相结合,利用专业知识进行剪枝,能够进一步减少候选项集数量,提高挖掘效率,提高故障预测准确率。将数据驱动与基于故障机理的模型驱动的故障预测方法相结合,改进算法,提高预测精度将是下一步的研究方向。

### 参考文献:

[1]陆啸秋,赵红卫,黄志平,等. 高速列车运行安全监控技术[J]. 铁道机车车辆,2011,31(2):34-37.  
LU Xiaoqiu,ZHAO Hongwei,HUANG Zhiping,et al. Safety monitoring technology of high speed train operation[J]. Railway Locomotive & Car,2011,31(2):34-37.

[2]黄学文,刘春明,冯璨,等. CHR3 高速动车组故障诊断系统[J]. 计算机集成制造系统,2010,16(10):2311-2318.  
HUANG Xuwen,LIU Chunwen,FENG Can,et al. Fault diagnosis system of CHR3 high speed EMUs[J]. Computer Inte-

表 11 频繁项目集 L5

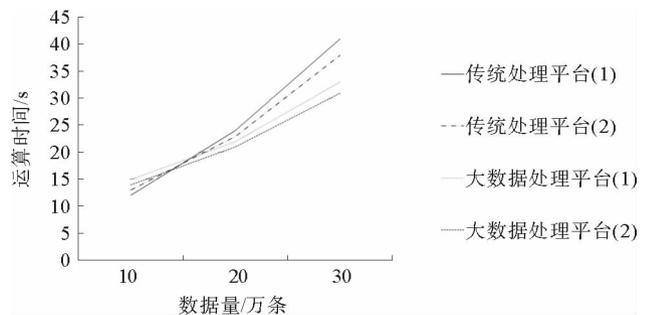
Tab. 11 Frequent item set L5

项目集	支持计数
{A1A2A3A4A6A7}	2 567

表 12 关联规则及其置信度

Tab. 12 Association rules and confidence coefficients

序号	规则	置信度/%
1	A1A2A3A4A6→A7	71.70
2	A1A2A3A4→A7	56.20
3	A1A2A3A6→A7	52.50
4	A2A3A4A6→A7	44.30
5	A1A3A6→A7	36.80
6	A2A3A6→A7	32.40



注:传统处理平台(1):在传统数据处理平台上采用无约束条件算法的运算时间;传统处理平台(2):在传统数据处理平台上采用约束条件算法的运算时间;大数据处理平台(1):在大数据处理平台上采用无约束条件算法的运算时间;大数据处理平台(2):在大数据处理平台上采用约束条件算法的运算时间

图 2 运算时间比较

Fig. 2 Comparison of running time

grated Manufacturing Systems,2010,16(10):2311-2318.

- [3]徐宇亮,孙际哲,陈西宏,等.基于加速退化试验和粒子滤波的电子设备故障预测方法[J].航空学报,2012,33(8):1483-1490.  
XU Yuliang,SUN Jizhe,CHEN Xihong,et al. Electronic equipment based on accelerated degradation test and particle filter [J]. Acta Aeronautica et Astronautica Sinica,2012,33(8):1483-1490.
- [4]曾庆虎,邱静,刘冠军.基于小波相关特征尺度熵的 HSMM 设备退化状态识别与故障预测方法[J].仪器仪表学报,2008,29(12):2559-2564.  
ZENG Qinghu, QIU Jing, LIU Guanjun. The method of HSMM equipment degradation state recognition and fault prediction based on wavelet correlation feature scale entropy[J]. Chinese Journal of Scientific Instrument,2008,29(12):2559-2564.
- [5]高甜容,于东,岳东峰,等.基于故障先兆判定模型和动态置信度匹配的主轴润滑故障预测方法[J].机械工程学报,2012,48(17):75-82.  
GAO Tianrong, YU Dong, YUE Dongfeng, et al. Prediction method of main shaft lubrication fault based on fault precursor judgment model and dynamic confidence matching[J]. Journal of Mechanical Engineering,2012,48(17):75-82.
- [6]KAI G, SAHA B, SAXENA A. A comparison of three data-driven techniques for prognostics[J]. IJET-IJENS. December 2010 IJENS I J E N S,2008:119-131.
- [7]BARALDI P, CADINI F, MANGILI F, et al. Model-based and data-driven prognostics under different available information [J]. Probabilistic Engineering Mechanics,2013,32(4):66-79.
- [8]WILSON Q, WANG, M. FARID G, et al. Prognosis of machine health condition using neuro-fuzzy systems[J]. Mechanical Systems and Signal Processing,2004,18(4):813-831.
- [9]AGRAWAL R, IMIELINSKI T, SWAMI A. Mining association rules between sets of items in large databases[C]//Proceedings ACM SIGMOD Conference on Management of Data, Washington D C, 1993,207-216.
- [10]AGRAWAL B R, SRIKANT R. A Fast algorithm for mining association rules[C]//中国粗糙集与软计算机学术研讨会,2003:619-624.
- [11]骆嘉伟,彭蔓蔓,陈景燕,等.基于消费行为的 Apriori 算法研究[J].计算机工程,2003,29(5):72-73.  
LUO Jiawei, PENG Manman, CHEN Jingyan, et al. Research on Apriori algorithm based on consumer behavior[J]. Computer Engineering,2003,29(5):72-73.
- [12]丛颖,刘其成,张伟.一种基于 Apriori 的微博推荐并行算法.计算机应用与软件,2015,32(8):229-233.  
CONG Ying, LIU Qicheng, ZHANG Wei. A parallel algorithm for micro-blog recommendation based on Apriori[J]. Computer Applications and Software,2015,32(8):229-233.
- [13]CHUNG S M, LUO C. Distributed mining of maximal frequent itemsets from databases on a cluster of workstations[C]//Cegrid, IEEE Computer Society,2004:499-507.
- [14]LUCCHESI C, ORLANDO S, PEREGO R. Fast and memory efficient mining of frequent closed itemsets[J]. IEEE Transactions on Knowledge & Data Engineering,2006,18(1):21-36.
- [15]LIU P Q, LI Z Z, ZHAO Y L. Effective algorithm of mining frequent itemsets for association rules[C]// International Conference on Machine Learning and Cybernetics, IEEE,2004:1447-1451.
- [16]O' DRISCOLL A, DAUGELAITE J, SLEATOR R D. 'Big data', Hadoop and cloud computing in genomics [J]. Journal of Biomedical Informatics,2013,46(5):774-781.

(责任编辑:傅 游)