

doi: 10.19562/j.chinasae.qcgc.2024.01.017

# 基于多维关联规则的车辆故障码解耦方法研究

胡杰<sup>1</sup>, 卿海华<sup>1</sup>, 魏敏<sup>2</sup>, 耿黄政<sup>2</sup>, 张潇<sup>1</sup>, 陈林<sup>1</sup>

(1. 武汉理工大学, 现代汽车零部件技术湖北省重点实验室, 现代汽车零部件技术湖北省协同创新中心, 新能源与智能网联车湖北工程技术研究中心, 武汉 430070; 2. 上汽通用五菱汽车股份有限公司, 柳州 545000)

**[摘要]** 本文提出一种针对车辆复杂耦合故障的故障码(DTC)解耦方法。首先由车辆故障自诊断原理及故障信号传播过程分析故障码的复杂关联性, 结合关联规则技术挖掘故障码间强关联关系, 并定义故障码多维关联规则; 其次由故障码数据集特征, 改进适用于故障码多维关联规则挖掘的FP-Growth算法; 最后由多维关联规则构建故障码关联知识图谱, 结合图论实现复杂故障码解耦。结果表明, 该方法能有效降低故障码的数量及复杂度, 提升基于故障码检修故障的效率。

**关键词:** 故障码; 解耦; 多维关联规则; 知识图谱

## Research on Vehicle Diagnostic Trouble Code Decoupling Method Based on Multidimensional Association Rule

Hu Jie<sup>1</sup>, Qing Haihua<sup>1</sup>, Wei Min<sup>2</sup>, Geng Huangzheng<sup>2</sup>, Zhang Xiao<sup>1</sup> & Chen Lin<sup>1</sup>

1. Wuhan University of Technology, Hubei Key Laboratory of Modern Auto Parts Technology, Hubei Collaborative Innovation Center for Automotive Components Technology, Hubei Research Center for New Energy & Intelligent Connected Vehicle, Wuhan 430070;  
2. SAIC General Wuling Automobile Co., Ltd., Liuzhou 545000

**[Abstract]** A DTC decoupling method for complex coupling faults of vehicles is proposed in this paper. Firstly, by analyzing the complex association of DTCs through the principle of vehicle fault self-diagnosis and the propagation process of fault signals, the strong association relationship between DTCs is mined combined with the association rule technology and the multidimensional association rules of DTC are defined. Secondly, the FP-Growth algorithm for DTC multidimensional association rule mining is improved by the characteristics of the DTCs dataset. Finally, the DTC association knowledge graph is constructed by multidimensional association rules to realize complex DTCs decoupling by combining graph theory. The results show that this method can effectively reduce the number and complexity of DTCs, and improve the efficiency of troubleshooting faults based on DTCs.

**Keywords:** diagnostic trouble code (DTC); decoupling; multidimensional association rule; knowledge graph

## 前言

随着车辆智能化程度的提升及电控模块数量的增加, 车辆产生电控故障的风险变大, 各电控模块间信号的密切关联也使得车辆在故障时会产生较多耦合故障码(diagnostic trouble code, DTC), 加大了基于故障码定位车辆故障原因的难度。因此, 解耦车辆

故障时产生的故障码, 有利于快速定位故障源头, 提升车辆故障检修效率。

车辆故障码是车辆发生故障时, 由电控单元(electronic control unit, ECU)记录并存储的一种故障编号, 作为反映车辆故障部件状态的关键信息, 故障码是国内外学者进行车辆故障诊断研究的重要数据基础。国外学者 Oliveira 等<sup>[1]</sup>基于二进制决策图(BDD)提出了一种利用主动和非主动 DTC 将多个

原稿收到日期为 2023 年 05 月 24 日, 修改稿收到日期为 2023 年 07 月 13 日。

通信作者: 胡杰, 教授, 博士生导师, E-mail: auto\_hj@163.com。

BDD组合成单个BDD的策略,并根据总诊断成本对策略进行了评估;Kriebe等<sup>[2]</sup>基于DTC产生时可能伴随的环境信息,提出了一种使用聚类技术并结合布尔逻辑、集合代数工具从DTC数据中进行故障模式识别的方法,但缺乏将识别结果与实际故障模式对比验证的环节;Thoorpu等<sup>[3]</sup>基于序列、非序列模型及频率分类技术构建了一种新的SDVE技术,将各种DTC信息进行矢量化,并构建深度样本学习模型,验证表明提升了零部件故障预测的准确性;Theissler等<sup>[4]</sup>基于汽车DTC及冻结帧数据,对汽车发生特定故障时汽车的运行模式进行检测研究,并成功地检测到汽车异常数据;Parivash等<sup>[5]</sup>基于DTC数据并利用随机森林从中导出其4类特征,作为故障预测性维护框架,把DTC作为不同部件即将发生故障的指标值。国内学者赵祥龙<sup>[6]</sup>基于车辆故障诊断案例等信息构建车辆故障知识图谱,融合XGBoost分类算法实现了汽车发动机故障原因分类,并实验验证了模型的准确性和有效性;王栩晨<sup>[7]</sup>总结三电系统故障机理和故障发生规律,制定了故障分类标准,以车辆监测数据为驱动,实现对三电系统部分故障至少提前10 s预测,且整体准确率在87%以上。国内外学者对车辆故障码的研究主要集中在故障码分类、模式识别、故障码序列预测及异常数据预警等方面,为车辆故障诊断提供了重要的研究支撑和技术指导,但当车辆故障产生大量故障码时,难以解耦复杂故障码集合、定位与故障底层原因密切相关的故障码。

故本文基于多维关联规则提出车辆故障码解耦方法,分析故障码的关联特性,结合关联规则挖掘技术将其复杂关联问题解耦为多维关联规则挖掘问题,改进适用于故障码多维关联规则挖掘的FP-Growth算法,构建故障码Neo4j关联知识图谱,并结合图论解耦复杂故障码,测试故障码的解耦效果。

## 1 故障码特性分析及应用

### 1.1 故障自诊断原理及传播分析

为方便描述,针对车辆故障时产生的故障码,作如下定义。源DTC:与故障部件直接相关、车辆产生故障时最先生成的DTC。衍生DTC:与故障部件间接相关、伴随源DTC生成的DTC。DTC解耦:从读取的故障码中去除衍生DTC,留下源DTC。

ECU故障自诊断原理如图1所示。前提条件即何时检测该信号,判断条件即判断该信号是否异常,

Debouncing算法通过阈值条件判断该异常是否为真实故障,故障存储策略即由ECU存储对应DTC,故若对异常信号的判断条件和阈值条件都不变,则发生同一故障时产生的故障码应是确定的。



图1 ECU故障自诊断系统原理

车辆故障信号传播过程如图2所示,发生单一故障时,伴随故障信号在各模块间传播,该故障部件所属模块会记录源DTC及衍生DTC,其他模块也可能记录衍生DTC,但引起这些故障码的底层故障原因应是一致的。经与某车企主机厂电控部门咨询确认,若多个电控模块同时发生故障,产生的故障码集合,其关联关系将更为复杂。故车辆故障时产生故障码集合,均存在源DTC。

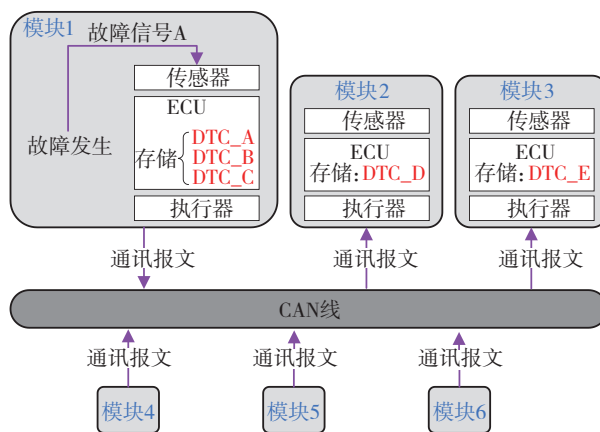


图2 车辆故障信号传播示意图

综上,车辆故障时,故障码的传播过程具有链式反应的特点,同时具有一定规律性,且都存在故障源头,即源DTC,故可以通过关联规则技术挖掘DTC间潜在的复杂关联性。

### 1.2 故障码关联分析

#### 1.2.1 关联规则挖掘技术

关联规则挖掘技术作为数据挖掘中常用的研究方法,用于提取和挖掘大型复杂数据中潜在的、未知的关联性知识<sup>[8]</sup>,多维关联规则是涉及多谓词或多属性的关联规则挖掘<sup>[9]</sup>。相关名词、挖掘步骤及相关性评价指标分别如下。

相关名词:项( $I$ ),项集( $S$ ),事务数据( $T$ ),数据库( $\mathcal{D}$ ),支持度,置信度,支持度阈值,置信度阈值。其中,支持度为含项 $I_i$ 的事务数据数量 $N_{I_i}$ 与事务数据总数量 $N_T$ 的比值,记为 $sup(I_i)$ ,计算公式见式(1)。置信度:同时含项集 $S_x$ 和 $S_y$ 的事务数据数量 $N_{S_x \cup S_y}$ 与含 $S_x$ 的事务数据数量 $N_{S_x}$ 的比值,记为 $conf(S_x \rightarrow S_y)$ ,计算公式见式(2)。支持度阈值:为消除较低频次项的负面影响,一般取0.05%。置信度阈值:关联规则可靠性的评价指标,本文结合实际情况取0.8(下文给出说明)。

$$sup(I_i) = \frac{N_{I_i}}{N_T} \quad (1)$$

$$conf(S_x \rightarrow S_y) = \frac{N_{S_x \cup S_y}}{N_{S_x}} \quad (2)$$

挖掘步骤:首先找到符合支持度阈值要求的项集,即频繁项集;其次计算频繁项集各项间置信度,符合置信度阈值要求的,即可构建多维关联规则<sup>[10]</sup>。

相关性评价指标:为修剪无效多维关联规则,需要设定相关性评价指标。常用评价指标有提升度、卡方系数、全置信度、最大置信度、Kulc、cosine、Leverage等,由于本文原始数据集预处理后不存在空项,且须同时计算频繁项集各项之间的置信度,故本文采取更方便的提升度作为多维关联规则相关性评价指标。提升度为置信度与支持度的比值,记为 $prom(S_x \uparrow S_y)$ ,如式(3)所示,由提升度定义知,其值大于1的正相关关联规则才有效<sup>[11]</sup>。

$$prom(S_x \uparrow S_y) = \frac{conf(S_x \rightarrow S_y)}{sup(S_y)} \quad (3)$$

### 1.2.2 故障码间强关联关系

运用多维关联规则挖掘技术挖掘车辆历史故障码数据中DTC的多维关联规则,相应有:项即DTC;项集即DTC集合;读取的车辆故障码集合即事务数据;所有历史故障码数据组成数据库 $\mathcal{D}$ 。

本文仅针对故障码的二维、三维关联规则进行挖掘,设有故障码A、B、C,则对应关联规则形式分别为:“ $A \rightarrow B$ 、 $A \cup B \rightarrow C$ ”,表示在置信度阈值条件下,前者发生时,后者也发生,即A与B、AUB与C为强关联关系。若故障码数据量足够多,由多维关联规则挖掘理论可知,在较高置信度要求下挖掘的多维关联规则是对原数据潜在的、未知知识的客观反映。一般来说,关联规则并不蕴含因果关系,但车辆故障码有其特殊性:一是DTC遵循车辆故障信号传播原理;二是DTC的释义赋予了DTC因果关系的特征。

所以故障码多维关联规则可认为是故障传播因果关系的揭露,即“ $A \rightarrow B$ ”、“ $A \cup B \rightarrow C$ ”分别揭露了“A是因、B是果”、“A、B是因、C是果”的关系<sup>[12]</sup>。

### 1.3 挖掘及应用方案

故障码多维关联规则挖掘及应用方案如图3所示,由车辆历史故障码数据集,数据预处理后得到有效故障码数据,改进FP-Growth算法挖掘故障码多维关联规则,通过Neo4j存储故障码多维关联规则知识,构建故障码关联知识图谱;输入车辆故障时诊断设备读取的DTC集合,解耦车辆故障码,输出源DTC集。

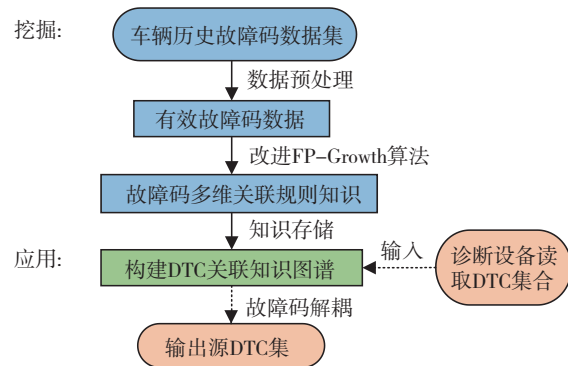


图3 故障码多维关联规则挖掘及应用方案

## 2 故障码多维关联规则挖掘

### 2.1 数据来源

本文原始数据为某企业新能源车型在2020年10月至2022年10月期间的故障车辆DTC清单,共计687 198条数据,初始格式如表1所示。

表1 故障车辆DTC清单初始格式

车型	Ecu	编码	状态	时间	Vin
E...	BMS	U180187	当前	2022/6/6 9:00:58	LK6...
E...	OBC	P1A2212	当前	2022/6/6 9:00:58	LK6...
E...	BMS	P180516	历史	2022/6/6 8:59:36	LK6...
E...	BMS	P181900	历史	2022/6/6 8:59:36	LK6...
E...	IC	U1C0287	当前	2021/7/2 11:41:55	LK6...
E...	ESP	C112224	当前	2021/7/2 11:41:55	LK6...
E...	...	...	...	...	...

### 2.2 数据预处理

由于原始数据存在“乱码、关键信息空值、故障码集合重复、含历史故障码、单次故障仅产生一个DTC”等问题,须按以下规则对数据预处理:去除乱

码、关键信息空值的无效数据;去除当前没有发生的历史状态故障码;去除同一车型同一时间重复上传的数据;去除仅含单一DTC的故障码数据。数据预处理后,得到26 294条有效事务数据,并以DTC所属模块分割,事务数据示例如表2所示,总计含21个模块,1 184个DTC。

表2 事务数据示例

事务数据	DTC所属模块			
	EPS	BMS	...	VCU
$T_1$	C100016, C166001, U198081	P110416, U007300	...	U1D0087, B108D00, B101914
$T_2$	U198087	P180616	...	B104A00, B102811
$T_3$	U140287, C166001	P182000, P110416	...	—
$T_4$	C166001	P182400	...	B100016, B104800
$T_5$	U198087, C166001	P180117, P182000, P182400	...	B102A00, U1A1181, B101914

### 2.3 FP-Growth算法改进

常用的频繁项集挖掘方法有FP-Growth、Apriori和Eclat算法。其中Apriori算法须多次扫描数据且会产生庞大的候选项集,计算量大;Eclat算法只适用于处理垂直格式的小型数据集;故本文选择更适合的FP-Growth算法挖掘频繁项集<sup>[13]</sup>。原始FP-Growth算法基于FP-Tree(频繁模式树)数据结构挖掘频繁项集。优点是只须扫描两次数据库,无须生成候选项集;缺点是数据量太大时,递归挖掘频繁项条件模式基,挖掘效率低,内存开销大<sup>[14]</sup>。针对此缺点,本文改进了FP-Growth算法中的FP-Tree,改进措施及改进后FP-Tree的构造步骤如下。

步骤1:初次扫描数据集,删除小于支持度阈值的项,对原始事务数据中各项按支持度降序排列,支持度相同的项,新增按哈希值降序排列。

步骤2:再次扫描数据集,按各项支持度大小由上往下降序创建项头表,用于存储各频繁项、其频次及其在FP-Tree的首节点位置,设定根节点Null,各事务数据由Null节点按特定规则向下生长<sup>[15]</sup>,并为所有节点新增nodeOrder属性,各节点nodeOrder值为它在项头表中的“逆序索引值”,根节点该值为项头表中项的总数,完成FP-Tree的构造。改进后FP-Tree节点属性及含义如表3所示。

表3 改进后FP-Tree节点属性及含义

节点属性	含义
Father	父节点
Children	子节点
Count	该分支节点频次
nodeLink	指向下一个分支相同节点
nodeOrder*	节点在项头表的逆序索引值*

注:带“\*”为新增的节点属性。

算法改进后,根据nodeOrder属性值过滤与项集频次计算无关的分支,理论上能提升挖掘效率。以同一计算机分别测试FP-Growth算法改进前后,挖掘相同事务数据样本相关频繁项集的耗时(样本量取2 000,4 000,6 000,8 000,10 000,20 000条),耗时结果统计如表4所示。由表可知:改进算法后,表中样本计算耗时明显降低,平均计算速度提升56.49%;且当数据量超过8 000条后,原生FP-Growth算法因内存溢出而无法输出结果,而算法改进后仍能较快输出结果。试验结果表明,算法改进后有效解决了原生FP-Growth算法挖掘大型数据频繁项集时内存溢出的问题,且能够提升对相关频繁项集的挖掘效率。

表4 算法改进前后挖掘所需频繁项集耗时

数据量	原算法 $t_1/s$	改进后 $t_2/s$	速度提升/%
2 000	6.74	2.25	66.70
4 000	7.91	3.94	50.18
6 000	8.23	4.76	42.16
8 000	16.01	5.30	66.90
10 000	内存溢出	11.43	
20 000	内存溢出	66.31	

注:耗时均在同一计算机运行得出。

### 2.4 挖掘流程

由FP-Tree构造步骤可知其有如下性质:树枝表示事务数据,节点的父节点必与该节点在同一事务数据中,且该节点Count属性值即该事务数据频次;分支中节点越靠近根节点,其Count、nodeOrder属性值越大,且其Count值等于其下一层所有子节点Count值之和;所有节点在各树枝中出现的先后顺序一致;节点可通过nodeLink属性链接下一分支中同名节点。

结合上述性质,并基于改进FP-Growth算法设计故障码多维关联规则挖掘方法,以节点A多维关联规则挖掘为例进行说明,其挖掘流程如图4所示(其中,X依次表示项头表中项A之前的项,Y代指X的父节点)。

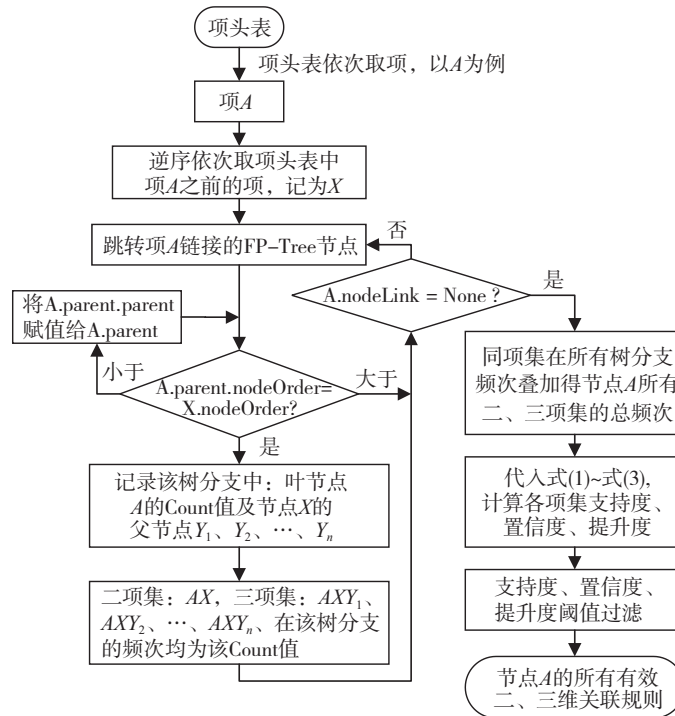


图4 节点A多维关联规则挖掘流程

设有数据集实例,其项头表及FP-Tree结构如图5所示(各节点Count、nodeOrder值已注明),分析节点A、D、E,易知  $N_A = 2, N_D = 5, N_E = 5, N_{A \cup D} = 2, N_{A \cup D \cup E} = 2$ ,代入式(1)~式(3)计算相应值,得

$$sup(A \rightarrow D) = \frac{2}{2} = 1$$

$$sup(D \rightarrow A) = \frac{2}{5} = 0.4$$

$$prom(A \uparrow D) = \frac{1}{5} = 0.2$$

$$sup(AD \rightarrow E) = \frac{2}{2} = 1$$

$$prom(AD \uparrow E) = \frac{1}{5} = 0.2$$

### 2.5 挖掘结果

输入预处理后的26 294条故障码事务数据和各

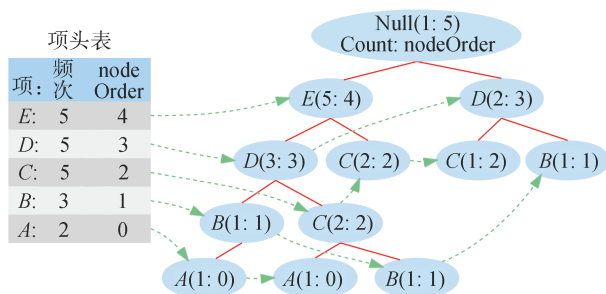


图5 项头表及FP-Tree结构实例

模块对应的故障码事务数据,根据上述故障码多维关联规则挖掘方法,输出故障码多维关联规则,去重后(如二维关联规则“ $A \rightarrow C$ ”和“ $B \rightarrow C$ ”组合等效于三维关联规则“ $A \cup B \rightarrow C$ ”)共计得有效故障码二维、三维关联规则各4 118、68 698条。

## 3 故障码知识图谱构建及解耦原理

### 3.1 故障码关联知识图谱构建

故障码关联规则数据量大且关系复杂,为方便查看、管理及运用故障码多维关联规则,采用能高效处理这类数据且能可视化展示关联规则的Neo4j知识图谱进行存储,作为结构化的语义知识库,其只有节点(实体)、边(关系)两类数据,且均可添加属性<sup>[16]</sup>。

Neo4j知识图谱的构建方式一般有3种:自顶向下、自底向上及两种方式相结合。由于本文研究的故障码为专业领域知识,故采用自顶向下的方式,其主要步骤是先定义知识图谱模式层,再从数据源中抽取实体、关系及其属性的结构化数据,与模式层相融合。根据原始故障码数据所蕴含的数据特征,定义故障码关联知识图谱模式层,如图6所示。上节挖掘的故障码多维关联规则为结构化数据,可直接导入数据层,由故障码多维关联规则分别构建故障

码节点的二维、三维关联关系,如故障码A、B有关联规则“ $A \rightarrow B$ ”,则可构建二维关联关系,由节点A指向节点B,三维关联规则同理。导入所有有效多维关联规则,得18 864个DTC节点,84 736条有向边(含二维、三维关联关系);同时,为方便进一步根据解耦后的源DTC辅助分析真实故障原因,由原始事务数据提取故障码频次、模块等数据,与故障码定义一起导入数据层,完成故障码关联知识图谱的构建,且通过Cypher查询语言可输出Neo4j知识图谱中存储的故障码节点关联情况及故障码所属模块信息,查询窗口及可视化界面如图7所示,故障码节点属性值

示例如表5所示。

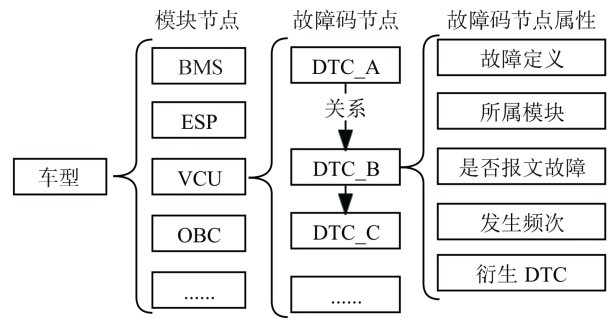


图6 故障码关联知识图谱模式层

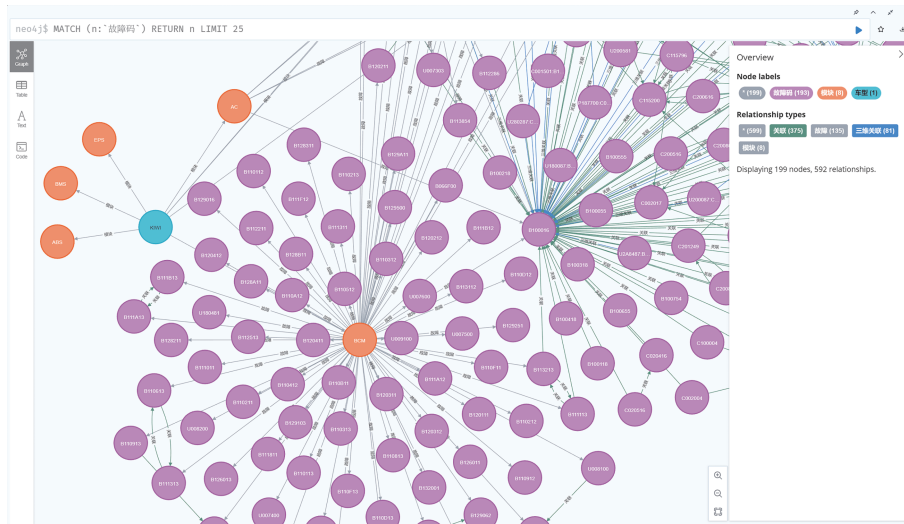


图7 Neo4j故障码关联知识图谱查询窗口及可视化界面

表5 故障码节点属性值示例

节点属性	属性值
<id>	412
name	B104B00
故障定义	电池包总电压状态——电压过低
所属模块	VCU
是否报文故障	否
发生频次	132
衍生DTC	B104A00、B104C00、B105C00、B107C00

### 3.2 故障码解耦原理

输入故障时读取的故障码,由故障码关联知识图谱查询DTC连通图,再由边的指向,即可定位源DTC,解耦故障码,其解耦原理如图8所示。由图论分析可知,定位源DTC的问题可转换为连通图中“入度为0”的节点或环的求解问题,由于存在节点成环情况,须先将连通图转化为有向无环图,才能通过“入度为0”筛选出所有源DTC节点,其求解步骤

如下。

步骤1:将强连通分量缩点,转化为有向无环图。如图8-I,首先,检测该连通图中的强连通分量,如图8-II中节点1、2、3,该分量任意两节点均有路径到达,彼此同时产生、同时消失;然后对所有强连通分量缩点,即用环内任意节点表示该强连通分量,得到有向无环图,如图8-III所示。

步骤2:输出“入度为0”的节点和环。在得到有向无环图后,便可直接根据“入度为0”条件,输出该连通图的所有源DTC。如图8-IV所示,缩点处理后,强连通分量(节点1、2、3)整体“入度为0”,故环内的节点均为源DTC。

为衡量、比较解耦效果,定义故障码精简率 $R$ ,如式(4)所示,其中, $N_1$ 表示解耦前DTC个数, $N_2$ 表示解耦后DTC个数, $R \in [0, 1)$ ,其值越大,故障码解耦效果越好。由于运用二维关联规则解耦效率更高,优先运用,得到的故障码解耦流程如图9所示。

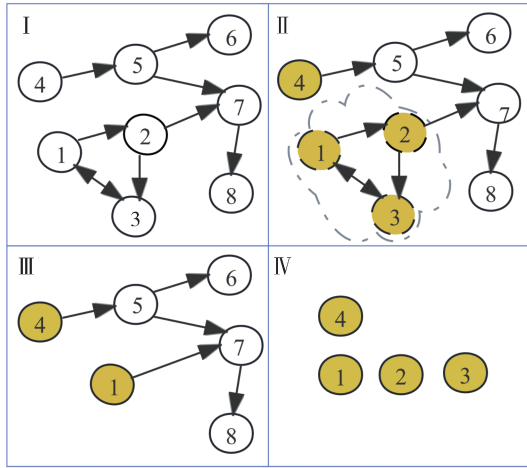


图8 故障码解耦原理

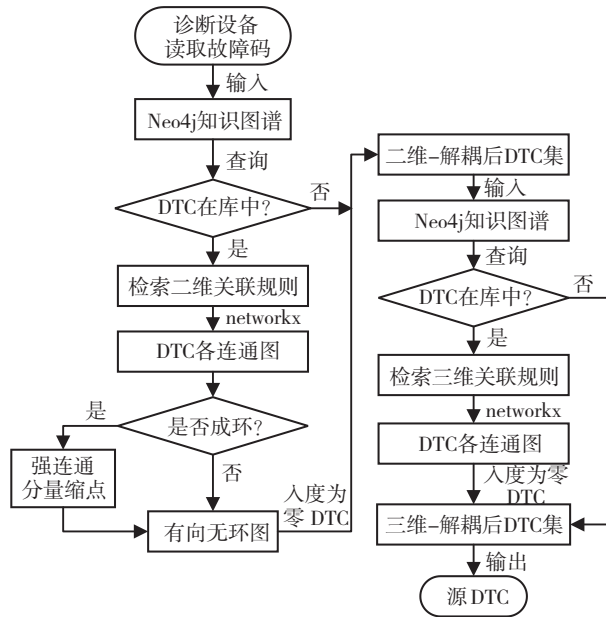
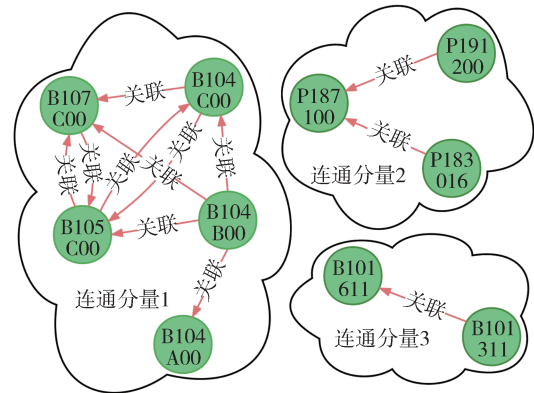


图9 故障码解耦流程

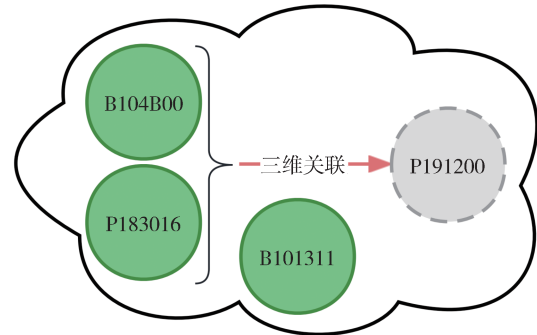
$$R = \frac{N_1 - N_2}{N_1} \times 100\% \quad (4)$$

如历史故障实例“P183016, P187100, P191200, B107C00, B104A00, B104C00, B105C00, B104B00, B101311, B101611”,其二维关联情况如图10(a)所示,由故障码解耦流程易知,二维解耦后DTC集为“B104B00, P91200, P183016, B101311”,该DTC集三维关联情况如图10(b)所示,易知三维解耦后,输出源DTC集为“P183016, B104B00, B101311”。

该故障实例相关DTC含义如表6所示,分析故障码含义可知,车辆报“B104B00:电池包总电压状态--电压过低”故障时,由于电源电压过低会引起“B104C00:电池内部CAN总线异常”、“B105C00:电



(a) 二维关联



(b) 三维关联

图10 故障实例

池包温度传感器异常”、“B107C00:电池包单体电压传感器异常”故障,与强连通分量“B107C00、B104C00、B105C00”同时产生、同时消失,且共同原因与“B104B00”的情况相符,总电压过低同样会引起“B104A00:电池组总压低于4级告警门限值”故障,都与图10(a)中二维关联情况相符;当电源电压过低且电池组总压低于4级告警门限值时,同样会引发“P191200:电池温度传感器故障”,与图10(b)中三维关联情况相符。

此时,再人工分析经多维关联规则解耦后的源DTC:“P183016, B104B00, B101311”,容易发现该故

表6 相关DTC含义

DTC	DTC含义
B104B00	电池包总电压状态--电压过低
B104C00	电池内部CAN总线异常
B105C00	电池包温度传感器异常
B107C00	电池包单体电压传感器异常
B104A00	电池包告警级别--电池停车请求
P183016	电池组总压低于4级告警门限值
P191200	电池温度传感器故障
B101311	PTC1使能回路对地短路

障真实原因是“PTC1使能回路对地短路(B101311)”,其直接导致电源电压过低、间接引起其它故障码。

### 4 故障码解耦测试及结果分析

基于故障码关联知识图谱,结合故障码解耦原理,解耦所有的历史故障码数据、高频次故障码数据及车辆常见故障的DTC集,并从不同层次计算故障码精简率,测试、分析故障码解耦效果。

考虑到高频次故障检修更频繁,应视为重要考虑因素,本文采用5-fold交叉验证法,计算在不同置信度阈值条件下,所有历史故障码事务数据、高频次历史故障码事务数据(定义频次≥5)的平均精简率,得到两类数据样本的平均精简率随置信度阈值变化的曲线,如图11所示,两种曲线变化趋势大致吻合,均呈两端式下降,前段下降较缓,后段下降较快,曲线交叉点在0.8附近,随着置信度阈值的提升,两类数据的平均精简率均下降越来越快,其主要原因是:阈值选取过大,一方面导致满足阈值条件的多维关联规则数量大幅减少从而直接引起精简率的下降,另一方面间接使强连通分量成环的可能性下降导致精简率下降。为使两类数据平均精简率均较优,取置信度阈值为0.8,此时两类数据平均精简率分别为55.62%、55.37%。

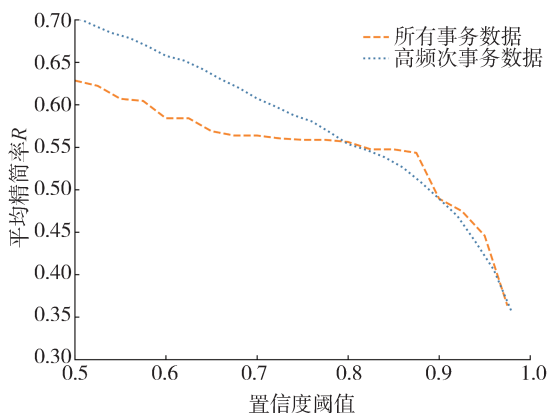
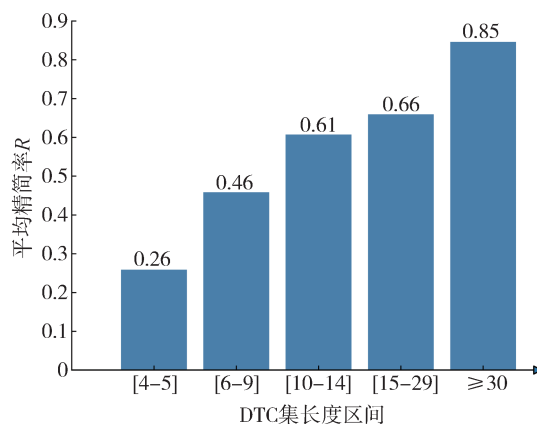
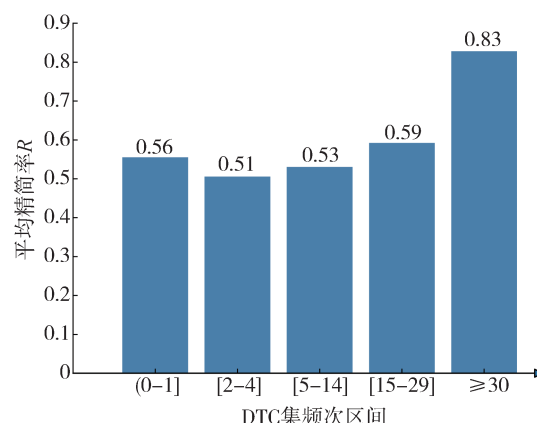


图11 所有数据、高频次数据各置信度阈值平均精简率

从DTC集的长度、频次范围,对比分析所有DTC集的解耦效果,结果如图12所示。由图12(a)可见,长度30及以上的DTC集,其精简率高达85%;由图12(b)可见,频次30及以上的DTC集精简率高达83%。选取车辆常见故障类型进行故障码解耦,结果如表7所示,所有常见故障的平均精简率为72.71%。



(a) 不同长度区间的平均精简率



(b) 不同频次区间的平均精简率

图12 不同层次DTC集解耦效果

表7 车辆常见故障的DTC集解耦效果

故障代号	常见故障	频次	精简率R/%
I	电池单体电压压差大	202	82.53
II	轮胎压力传感器故障	181	74.15
III	左前轮速传感器故障	156	69.51
IV	充电模块故障	57	58.13
V	电池单体温度过高	109	79.24

### 5 结论

本文从车辆ECU故障自诊断原理及电控模块内外信号的传播过程出发,结合关联规则挖掘技术分析故障码的复杂关联性,定义了故障码多维关联规则并通过改进FP-Growth算法挖掘,构建故障码关联知识图谱,结合图论解耦车辆故障码。以文中数据为样本输入,该故障码解耦方法平均能去除55.62%的衍生DTC,长度、频次分别在30及以上的

DTC集平均精简率高达85%、83%,对常见故障DTC集的平均精简率为72.71%。实验结果表明,基于多维关联规则的车辆故障码解耦方法能有效降低故障码的数量及复杂度,提升基于故障码检修车辆故障的效率,辅助维修人员快速分析车辆故障底层原因。

### 参考文献

- [1] OLIVEIRA M F, LUDERS R. Combining multiple diagnostic trouble codes into a single decision tree [J]. IFAC PAPERSONLINE, 2016, 49(11): 555-561.
- [2] KRIEBE S, KUSMENKO E, RUMPE B, et al. Learning error patterns from diagnosis trouble codes [C]. 2019 IEEE Intelligent Vehicles Symposium (IV), 2019: 179-184.
- [3] THOORPU K R, PRAFULLA N. Sequential DTC vector embedding using deep neural networks for industry 4.0 [C]. 2020 IEEE 7th International Conference on Industrial Engineering and Applications (ICIEA), 2020: 912-915.
- [4] THEISSLER A. Multi-class novelty detection in diagnostic trouble codes from repair shops [C]. 2017 IEEE 15th International Conference on Industrial Informatics (INDIN), 2017: 1043-1049.
- [5] PIRASTEH P, NOWACZYK S, PASHAMI S, et al. Interactive feature extraction for diagnostic trouble codes in predictive maintenance [P]. Interactive Data Mining, 2019.
- [6] 赵祥龙. 基于第三方云平台的车辆故障知识图谱构建 [D]. 成都:西南交通大学, 2019.  
ZHAO X L. Vehicle fault knowledge graph based on third-party cloud platform [D]. Chengdu: Southwest Jiaotong University, 2019.
- [7] 王栩晨. 基于数据驱动的电动汽车三电系统故障诊断及预测研究 [D]. 西安:长安大学, 2022.  
WANG X C. Research on fault diagnosis and prediction of electric vehicle three electric systems based on data driven [D]. Xi'an: Chang'an University, 2022.
- [8] YAMAMOTO H, KAYANUMA G, NAGASHIMA T, et al. Early detection of adverse drug reaction signals by association rule mining using large-scale administrative claims data [J]. Drug Safety, 2023.
- [9] 杨勇. 多维关联规则挖掘在烟草行业中的应用探索 [J]. 科技创新与应用, 2022, 12(12): 55-58.  
YANG Y. Application exploration of multi-dimensional association rule mining in tobacco industry [J]. Scientific and Technological Innovation and Application, 2022, 12(12): 55-58.
- [10] 徐凯. 基于关联规则的电商大数据挖掘与应用研究 [D]. 西安:西京学院, 2022.  
XU K. Research on e-commerce big data mining and application based on association rules [D]. Xi'an: Xijing University, 2022.
- [11] DENG N, YE Z W, CAI W H, et al. Association rule mining based on hybrid whale optimization algorithm [J]. International Journal of Data Warehousing and Mining (IJDWM), 2022, 18(1).
- [12] 柳媛. 基于无监督学习的故障分析研究 [D]. 北京:北京邮电大学, 2021.  
LIU Y. Failure analysis research based on unsupervised learning [D]. Beijing: Beijing University of Posts and Telecommunications, 2021.
- [13] ARIVAZHAGAN B, SUBRAMANIAN R S. Hybrid model for pattern discovery in data communication to enhance customer relationship management using data mining techniques [J]. Journal of Physics: Conference Series, 2021, 1979(1).
- [14] SHAWKAT M, BADAWI M, EI-GHAMRAWY S, et al. An optimized FP-Growth algorithm for discovery of association rules [J]. The Journal of Supercomputing, 2021, 78(4).
- [15] 王营, 高琦, 李婷玉, 等. 基于改进FP-growth算法的售后服务数据挖掘 [J]. 现代制造工程, 2021(6): 31-37.  
WANG Y, GAO Q, LI T Y, et al. After-sales service data mining based on improved FP-growth algorithm [J]. Modern Manufacturing Engineering, 2021(6): 31-37.
- [16] 周刚, 闫龙川, 吴小华, 等. 面向电力传输信息系统故障定位的知识图谱构建与应用 [J]. 激光杂志, 2021, 42(12): 190-196.  
ZHOU G, YAN L C, WU X H, et al. Construction and application of knowledge graph for fault location of power transmission information system [J]. Laser Magazine, 2021, 42(12): 190-196.