

# 基于特征信息与故障识别的知识图谱构建方法

杨伊悦<sup>1</sup>, 郭奇<sup>2</sup>, 蒋磊<sup>1</sup>, 才进<sup>1</sup>, 杜国<sup>1</sup>, 赵月静<sup>1</sup>, 秦志英<sup>1\*</sup>

1. 河北科技大学 机械工程学院, 河北石家庄 050018

2. 金叶天成(北京)科技有限公司 研发中心, 北京 100024

**摘要:** 为提高工业设备故障诊断的准确性与可解释性, 提出了一种基于特征信息与故障识别的知识图谱构建方法。通过采集渣浆泵振动信号, 利用最优类距比对时域、频域以及熵特征进行降维, 筛选出敏感特征参数, 同时结合多源信号融合与 DATSDP 图像变换生成图像特征, 并在此基础上定义多层级的知识图谱实体类型并完成实体及属性抽取。通过 SVM 和 AlexNet 模型分别实现参数与图像特征的故障分类, 建立实体间关系链, 完成关系抽取。利用 Cypher 语句将知识三元组导入 Neo4j 图数据库, 构建出了包含 36 010 个实体节点和 39 609 条关系的知识图谱。实验表明, 参数特征结合 SVM 分类器平均准确率达 99.67%, 图像特征在 AlexNet 模型中准确率达 99.72%, 验证了方法的有效性和鲁棒性。该方法为设备状态监测与故障诊断提供了结构化知识表达与高效检索支持。

**关键词:** 参数特征; 图像特征; 故障识别; 实体及属性抽取; 关系抽取; 知识图谱

**文献标志码:** A **中图分类号:** TP391.4

## Construction method of knowledge graph based on feature information and fault identification

YANG Yiyue<sup>1</sup>, GUO Qi<sup>2</sup>, JIANG Lei<sup>1</sup>, CAI Jin<sup>1</sup>, DU Guo<sup>1</sup>, ZHAO Yuejing<sup>1</sup>, QIN Zhiying<sup>1</sup>

1. School of Mechanical Engineering, Hebei University of Science and Technology, Shijiazhuang 050018, China

2. Center of Research and Development, KINGYEE(Beijing) Co., LTD., Beijing 100024, China.

**Abstract:** In order to improve the accuracy and interpretability of industrial equipment fault diagnosis, a knowledge graph construction method based on feature information and fault identification is proposed. By collecting vibration signals of slurry pump, the dimension of time domain, frequency domain and entropy features are reduced by using optimal class distance ratio, and sensitive feature parameters are selected. At the same time, image features are generated by combining multi-source signal fusion and DATSDP image transformation, and on this basis, multi-level knowledge graph entity types are defined and entity and attribute extraction is completed. The fault classification of parameters and image features is realized by SVM and AlexNet model respectively, and the relationship chain between entities is established to complete the relationship extraction. Using Cypher statement, the knowledge triplet was imported into the secondary graph database, and a knowledge graph containing 36 010 entity nodes and 39 609 relationships was constructed. Experiments show that the average accuracy of parameter features combined with SVM classifier is 99.67%, and the accuracy of image features in AlexNet model is 99.72%, which verifies the effectiveness and robustness of the method. This method provides structured knowledge expression and efficient retrieval support for equipment condition monitoring and fault diagnosis.

**基金项目:** 国家自然科学基金项目 (12402016)。

**作者简介:** 杨伊悦 (2001—), 男, 硕士研究生, 研究方向为知识图谱、故障诊断; 秦志英 (1976—), 通信作者, 女, 博士, 教授, 研究方向为智能运维与故障诊断, E-mail: qinzhy76@163.com

**收稿日期:** 2025-00-00 **修回日期:** 2025-00-00

**Key words:** parameter characteristics; image features; fault identification; entity and attribute extraction; relationship extraction; knowledge graph

随着工业化进程的加速,机械系统的复杂性和运行环境的不确定性显著增加,这对设备的安全性与稳定性提出了更高要求。工业设备故障不仅可能导致生产中断、经济损失,甚至引发重大安全事故<sup>[1]</sup>。因此,高效故障诊断技术成为保障生产安全的核心需求。

近年来,随著计算机技术的快速发展,为机械设备的故障诊断提供了新思路。目前主流方法可分为两类:基于参数特征的机器学习方法和基于图像特征的深度学习方法。机器学习方法通过时域、频域分析提取振动信号特征,并输入随机森林、支持向量机(SVM)等模型进行分类。例如,叶韬等人<sup>[2]</sup>通过优化SVM参数并提取离心泵振动信号的时域特征,诊断准确率在90%以上;Wang等人<sup>[3]</sup>结合集合经验模态分解和相关系数阈值去噪技术提取特征频率,有效识别机械部件的故障模式;胡泽等人<sup>[4]</sup>利用Hilbert-Huang变换对振动信号提取特征,输入到神经网络进行模式识别,实现对离心泵状态的故障识别。深度学习作为机器学习的重要分支,能够自动进行特征提取,其中卷积神经网络(CNN)因其局部感受野、权值共享和池化的优势而广泛应用,且更适合处理图像数据。庞新宇等人<sup>[5]</sup>采用GAF方法将振动信号转换为二维图像并通过CNN实现高精度诊断;侯玉叶等人<sup>[6]</sup>则通过SDP方法缩短计算时间并提升CNN模型故障诊断效率。故障诊断算法在工业领域中展现出了一定的优势,但在实际应用中单一利用设备机理知识或设备运行数据,难以解决多复杂工况、多故障类型的设备故障问题。

知识图谱作为一种新兴的信息处理技术<sup>[7]</sup>,通过结构化表示实体之间的关系,挖掘和表达复杂的关联信息。根据应用目标可以将其划分为通用领域知识图谱和垂直领域知识图谱<sup>[8]</sup>。通用领域知识图谱具有极为广泛的信息类别,包括从科学、文化到日常生活中的各种知识,例如,中科院计算机语言信息工程研究中心的HowNet模型<sup>[9]</sup>和上海交通大学Zhishi.me模型<sup>[10]</sup>等,这类知识图谱能够丰富中文信息资源,为文本理解、数据分析以及智能问答等提供坚实基础。

垂直领域知识图谱<sup>[11]</sup>则更专注于特定行业的专业知识。在医疗领域,知识图谱能够包含详细的疾病信息、治疗方法及最新的研究成果,HUANG等人<sup>[12]</sup>通过整合多种类型的医疗数据并使用多模态预训练模型,构建出多模态知识图谱,并将其应用于医疗方面以提升诊断和治疗的效果。在教育领域,知识图谱能

够进行结构化知识表达、个性化学习支持以及智能化教学辅助,JIANG等人<sup>[13]</sup>采用教育本体来建模在线学习资源的知识结构,构建了与MOOC相关的知识图谱,为学者提供更加个性化和高效的学习体验。在工业领域,知识图谱能够通过结构化知识建模、多源数据融合与智能推理,实现生产流程优化和故障智能诊断,YANG等人<sup>[14]</sup>提出了一种本体引导的多层次知识图谱,能够提供良好的特征表示,在高炉炼铁过程中得到了有效应用;唐荻音<sup>[15]</sup>提出一种本体与实体双向约束的知识图谱构建方法,解决了航天器故障形式多样、差异大、数量少以及遥测数据所蕴含知识难以被有效利用的缺陷;陈宏<sup>[16]</sup>提出一种基于知识图谱的风力发电机故障诊断系统,解决了故障诊断与维修过程不明确以及历史故障数据记录大量遗留等问题;刘晶<sup>[17]</sup>等人提出了一种基于知识和数据融合驱动的设备故障诊断方法,从单纯依赖机理知识或运行数据到两者融合驱动,进一步形成故障图谱诊断系统。

尽管知识图谱技术在医疗、金融等领域的应用已取得了一定成果,但在工业故障诊断领域,尤其是针对振动信号等非文本数据的应用较少,同时面临着诸多挑战。目前研究表明,已有方法在特定条件下能够取得较好的效果,但面对多复杂工况、多故障类型以及数据关联的设备故障问题时,单一依靠设备机理知识或设备运行数据难以提供全面有效的解决方案。

## 1 基于特征信息与故障识别的知识图谱构建流程

基于以上分析,本文结合故障诊断与知识图谱技术,提出了一种基于特征信息与故障识别的知识图谱构建方法。首先,收集实验设备的振动信号,依据数据集结构定义知识图谱的层次化实体类型,并进行实体抽取;其次,提取时域参数、频域参数和熵参数,并通过最优类距比对参数进行筛选得到敏感特征;对多源信号整合并通过DATSDP变换得到图像特征;提取出两部分特征作为相应实体节点的内部信息,完成属性抽取;之后,将两种特征分别输入到诊断算法中进行分类训练,并根据分类结果完成实体之间的关系抽取;最后,利用Cypher语句将抽取到的知识三元组转换成图数据库存储格式,生成实体以及实体之间的关系,进而构建知识图谱。

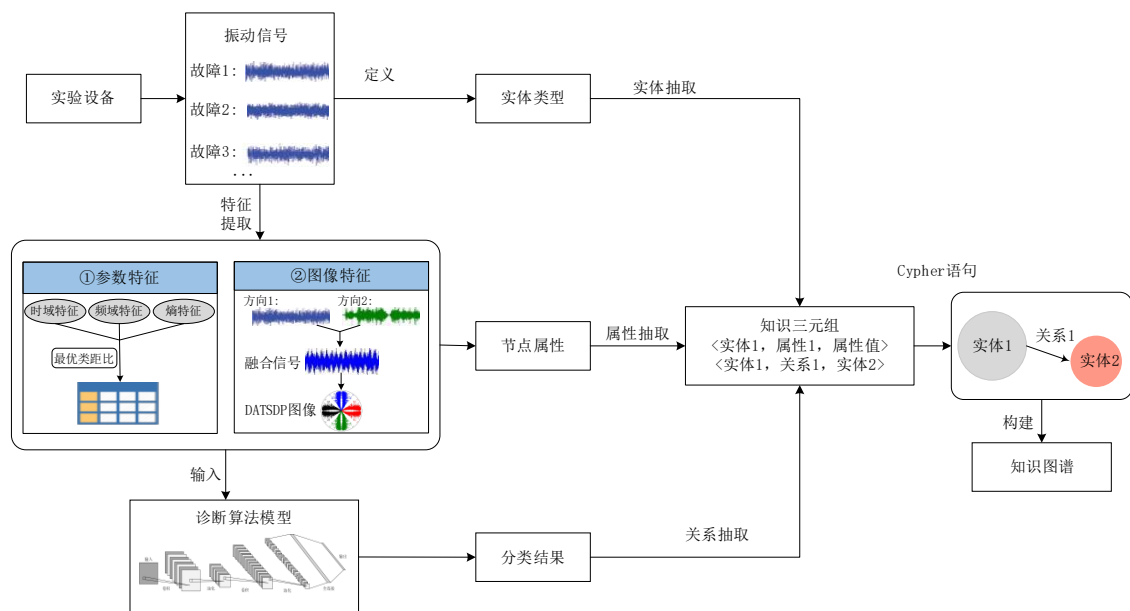


图1 特征图谱的构建流程

Fig.1 Construction process of characteristic graph

## 2 知识图谱的实体抽取

实体抽取也称为命名实体识别,是指从非结构化或半结构化文本等数据中识别出特定类型的实体,并将这些实体分类到预定义实体类型中的过程。本文使用的数据为振动信号数据,此类数据具有时序性、多维特征和物理意义明确的特点。与传统文本数据不同,振动信号中的实体并非人名、地名等概念,而是需要根据数据结构和机理知识进行特殊实体类型的定义,进而完成实体抽取。

### 2.1 振动信号数据的采集与处理

以 200ZJ-A60 型号渣浆泵为实验对象,采集其加速度振动信号用于后续研究。实验模拟了三种设备状态包括,拧松地基上的螺栓设置地基松动故障,调整泵轴与叶轮的安装角度设置泵轴不平衡故障,以及设备健康状态。设置如图 2 所示的两个测点,使用三轴加速度传感器和 DH5902 坚固型数据采集设备及配套软件,进行数据采集与存储,采样频率设为 100 000 Hz。

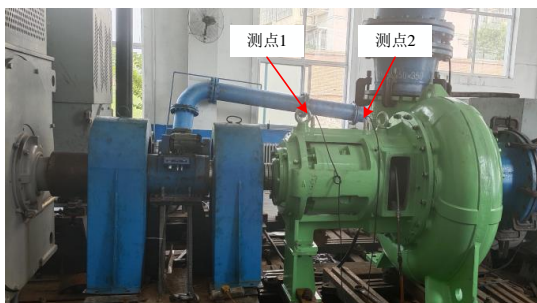


图2 200ZJ-A60 型号渣浆泵

Fig.2 200ZJ-A60 slurry pump

提取出其中两个测点的 x、y 方向信号,连续 5 120 个数据点作为一个样本,每种设备状态包含 300 个样本,并按照 8: 2 的比例划分训练集及测试集,其中表 1 展示了实验数据划分的详细信息。

表1 实验数据划分信息

Table 1 Experimental data division information

信号类型	训练/测试样本数量	标签
测点 1 健康状态 (health1)	240/60	0
测点 2 健康状态 (health2)	240/60	1
测点 1 不平衡故障 (imbalance1)	240/60	2
测点 2 不平衡故障 (imbalance2)	240/60	3
测点 1 地基松动 (loose1)	240/60	4
测点 2 地基松动 (loose2)	240/60	5

### 2.2 实体抽取

根据划分后数据集的数据结构和相关信息进行总结归纳,定义出知识图谱中实体的具体类型。划分后的数据包含了测点 1 和测点 2 的 x、y 两个方向上的振动信号,对其进行总结分类,得到的具体结果如表 2 所示。

表2 数据结构与信息总结表

Table 2 Data structure and information summary table

数据信息	描述	分类指标
测点位置	传感器的安装位置	测点 1、测点 2
采集方向	振动信号的不同采集方向	x 方向、y 方向
故障类型	数据集包含的设备状态	健康状态、地基松动、不平衡状态
故障特征	从数据集种提取的特征	参数特征, DATSDP 融合图像

其中分类指标包括了振动信号数据的测点位置、采集方向、故障类型和故障特征 4 类。根据以上规则进行实体类型划分, 并进行实体级别的设置, 结果如表 3 所示。其中六级实体类型包含参数与图像特征; 五级实体类型包含对应的 x、y 方向信号; 四级实体类型为 x、y 两方向信号信息汇总后的综合实体节点; 三级实体类型包含六种故障分类; 二级实体类型包含三种设备状态; 一级实体类型则为振动信号数据来源。另外, 划分的实体类型同时作为知识图谱中实体节点的名称 (name), 进而完成实体抽取。

表 3 实体类型名称表

Table 3 Entity type name table

实体类型等级	实体类型
一级 (一种实体类型)	渣浆泵信号数据集 健康状态
二级 (两种实体类型)	地基松动 泵轴不平衡 测点 1 健康状态信号
三级 (六种实体类型)	... 测点 2 不平衡故障信号
四级 (一种实体类型)	xy 方向信号
五级 (两种实体类型)	x 方向信号 y 方向信号 x 峰峰值
六级 (十七种实体类型)	... DATSDP 融合图像

### 3 知识图谱的属性抽取

属性抽取通常是指从非结构化的文本数据中识别并提取出特定实体的属性信息。具体而言, 是从大量的自然语言文本中自动地找到关于某个对象(如人物、地点、组织等)的具体特征或描述, 并将这些信息以结构化形式进行表示。然而, 对于振动信号数据这种非文本信息, 并不具有可读的特征描述, 因此需要采用特征提取方法来获得相关的参数特征或图像特征, 来间接描述振动数据所包含的信息。

#### 3.1 参数特征提取

##### 3.1.1 时域特征

时域特征参数<sup>[18]</sup>根据是否具有物理单位, 可分为有量纲参数和无量纲参数。有量纲参数提供关于信号绝对强度和能量的具体信息, 适用于评估系统的绝对

性能或状态; 而无量纲参数则反映了信号内部结构和形态的变化, 适合用于比较不同信号之间的相对特性。具体计算公式如表 4、表 5 所示。表中  $x(k)$  是时域信号序列,  $k=1,2,3,\dots,N$ ,  $N$  为样本点数。

表 4 有量纲时域特征及计算公式

Table 4 Dimensional time domain characteristics and calculation formulas

特征名称	计算公式	特征名称	计算公式
均值	$F1 = \frac{1}{N} \sum_{k=1}^N x(k)$	峰峰值	$F2 = \max(x(k)) - \min(x(k))$
标准差	$F3 = \sqrt{\frac{\sum_{k=1}^N (x(k) - F1)^2}{N - 1}}$	均方根值	$F4 = \sqrt{\frac{1}{N} \sum_{k=1}^N (x(k))^2}$

表 5 无量纲时域特征及计算公式

Table 5 Dimensionless time domain characteristics and calculation formulas

特征名称	计算公式	特征名称	计算公式
峭度	$F5 = \frac{\frac{1}{N} \sum_{k=1}^N (x(k) - F1)^4}{\left(\frac{1}{N} \sum_{k=1}^N (x(k) - F1)^2\right)^2}$	偏度	$F6 = \frac{\frac{1}{N} \sum_{k=1}^N (x(k) - F1)^3}{\left(\frac{1}{N} \sum_{k=1}^N (x(k) - F1)^2\right)^{3/2}}$
峰值指标	$F7 = \frac{\max x(k) }{F4}$	波形指标	$F8 = \frac{F4}{\frac{1}{N} \sum_{k=1}^N  x(k) }$

##### 3.1.2 频域特征

频域特征<sup>[19]</sup>则能够更深入地理解信号的本质特征, 还能揭示信号中的周期性成分以及噪声水平。计算公式如表 6 所示, 表中  $s(k)$  是对应功率谱值,  $f_k$  是对应点的频率幅值大小,  $k=1,2,3,\dots,N$ ,  $N$  为样本点数。

表 6 频域特征及计算公式

Table 6 Frequency domain characteristics and calculation formula

特征名称	计算公式	特征名称	计算公式
重心频率	$F9 = \frac{\sum_{k=1}^N f_k \cdot s(k)}{s(k)}$	平均频率	$F10 = \frac{\sum_{k=1}^N s(k)}{N}$
均方根频率	$F11 = \sqrt{\frac{\sum_{k=1}^N f_k^2 \cdot s(k)}{\sum_{k=1}^N s(k)}}$	频率标准差	$F12 = \sqrt{\frac{\sum_{k=1}^N (f_k - F9)^2 \cdot s(k)}{\sum_{k=1}^N s(k)}}$

##### 3.1.3 熵特征

熵<sup>[20]</sup>则用于量化系统的无序程度或不确定性。在信号处理和数据分析中, 熵特征被广泛应用于描述信号的复杂性、随机性和信息含量。计算公式如表 7 所

示,表中  $p(x(k))$  为信号取值为  $x(k)$  的概率;  $m$  为嵌入维度;  $r$  为距离容忍度;  $C_k^m(r)$  为匹配概率,每个长度为  $m$  的子向量  $X_k^m$ ,与其他子向量  $X_l^m$  之间的距离  $d[X_k^m, X_l^m]$ ,不大于  $r$  的匹配概率;  $\ln$  为以  $e$  为底的对数;  $B_k^m(r)$  同样为匹配概率,但其中  $k \neq l$ ;  $\mu(d[X_k^m, X_l^m])$  为模糊相似度。

表 7 熵特征及计算公式

Table 7 Entropy characteristics and calculation formula

特征名称	计算公式
信息熵	$F13 = -\sum_{k=1}^N p(x(k)) \cdot \log_2 p(x(k))$
近似熵	$F14 = \frac{\sum_{k=1}^{N-m+1} \ln(C_k^m(r))}{N-m+1}$
样本熵	$F15 = -\ln \left( \frac{\sum_{k=1}^{N-m} B_k^{m+1}(r)}{N-m} \right) / \left( \frac{\sum_{k=1}^{N-m+1} B_k^m(r)}{N-m+1} \right)$
模糊熵	$F16 = -\ln \left( \frac{\sum_{k=1}^{N-m+1} \left( \frac{1}{N-m} \sum_{l=1, l \neq k}^{N-m+1} \mu(d[X_k^{m+1}, X_l^{m+1}]) \right)}{\sum_{k=1}^{N-m+1} \left( \frac{1}{N-m} \sum_{l=1, l \neq k}^{N-m+1} \mu(d[X_k^m, X_l^m]) \right)} \right)$

### 3.2 基于最优类距比的参数特征降维

上述参数虽然在一定程度上反映了系统状态或样本特性,但存在着计算复杂度高、模型泛化能力下降等问题。因此,需要有效地提取具有代表性的特征参数,降低数据维度,以提升故障识别效率,提升故障分类模型的效率和性能。

最优类距比<sup>[21]</sup>用于描述不同类别之间最优分离程度,尤其是在聚类分析和监督学习中。该指标能够判断不同类别的数据是否有效区分,同时保持同一类别数据内的紧凑性。对于给定的数据集  $D$  和其对应的类别标签  $C$ ,计算过程为:

#### 1. 计算类别中心 $\mu_k$

计算所有样本在每个特征上的平均值,作为每个类别的中心,计算公式为:

$$\mu_k = \frac{1}{N_k} \sum_{x_i \in C_k} x_i \quad (1)$$

其中,  $N_k$  为第  $k$  类的样本数;  $C_k$  为属于第  $k$  类的所有样本点。

#### 2. 计算类内距离 $D_{intra}$

类内距离是指同一类别内所有样本点到该类别中心的平均距离,计算公式为:

$$D_{intra}(k) = \frac{1}{N_k} \sum_{x_i \in C_k} \|x_i - \mu_k\|^2 \quad (2)$$

然后对所有类别进行均值处理,得到总体类内距离:

$$D_{intra} = \frac{1}{K} \sum_{k=1}^K D_{intra}(k) \quad (3)$$

其中,  $K$  为总类别数。

#### 3. 计算类间距离 $D_{inter}$

类内距离是指不同类别中心之间的平均距离或最短距离,其计算公式为:

$$D_{inter} = \frac{1}{K(K-1)} \sum_{i=1}^K \sum_{j \neq i} \|\mu_i - \mu_j\|^2 \quad (4)$$

#### 4. 计算最优类距比 $R$

根据类间距离与类内距离,计算每个特征的最优类距比,并对其进行排序,筛选出分类效果更优的特征参数。其中最优类距比的计算公式为:

$$R = \frac{D_{inter}}{D_{intra} + \varepsilon} \quad (5)$$

其中,  $\varepsilon$  为防止分母为零,设置的数值极小的常数。

根据特征计算公式提取相关特征,通过最优类距比分别筛选出排名靠前的参数特征,将这些特征进行降维,得到最终的与信号密切相关的特征参数。以每个工况的  $x$  方向数据集为输入,根据以上有量纲时域特征计算公式,对五种工况下的训练集数据进行时域特征提取,最终提取到的部分时域特征值如表 7 所示。

表 8 部分信号的有量纲时域特征值 ( $\times 10^{-3}$ )Table 8 Dimensional time domain eigenvalues of some signals ( $\times 10^{-3}$ )

工况	均值	峰峰值	标准差	均方根值	标签
测点 1 健康状态	1.29	98.56	9.08	9.17	0
	1.27	80.51	9.04	9.13	0
测点 2 健康状态	1.31	97.57	5.73	5.75	1
	1.33	81.36	5.70	5.72	1
测点 1 不平衡故障	0.85	30.73	4.35	4.43	2
	0.86	34.54	4.44	4.52	2
测点 2 不平衡故障	0.91	96.61	9.70	9.74	3
	0.88	95.13	9.82	9.86	3
测点 1 地基松动	0.91	44.76	5.25	5.33	4
	0.91	43.56	5.07	5.15	4
测点 2 地基松动	0.82	69.48	8.47	8.51	5
	0.83	69.48	8.67	8.71	5

之后,利用最优类距比对特征进一步挖掘,筛选出分类效果相对较优的特征参数,具体结果如表 8 所示。由表中最优类距比排序,筛选出峰峰值与标准差,以此对特征参数进行降维得到二维特征,作为后续的特征输入。





然后,将特征参数和 DATSDP 融合图像等信息,作为四级、五级和六级实体节点的特征属性。例如:“x 峰峰值”实体节点需要输入 x 方向信号的峰峰值;“DATSDP 融合图像 1”实体节点需要输入融合图像的 URL 地址与图像数据;“x 方向信号”实体节点需要输入 x 方向信号的振动数据、对应参数特征向量以及图像实体节点中的信息;“xy 方向信号”实体节点则需要汇总对应时间段 x、y 方向信号包含的全部信息。下面列举出部分实体的具体属性内容,如表 9 所示。

表 10 部分实体节点及属性内容

Table 10 Some Entity Nodes and Attribute Contents

实体等级	实体名称	实体属性
六级	x 峰峰值	value: 0.03751
六级	DATSDP 融合图像	DATSDP_data_0:[[0.3602,0.4368,...,0.2061]] image_URL:'file:///.../DATSDP_1.jpg'
五级	x 方向信号	x_signal:[-2.1561,3.8574,5.9384,...,-1.0178] feature:[0.03751,0.04415,...,0.06446][1×8] DATSDP_data_0:[[0.3602,0.4368,...,0.2061]] image_URL:'file:///.../DATSDP_1.jpg'
四级	xy 方向信号	x_signal:[-2.1561,3.8574,5.9384,...,-1.0178] y_signal:[4.21819,-1.5473,0.3351,...,4.2157] xy_feature:[0.03751,0.04415,...,5.2234][1×16] DATSDP_data_0:[[0.3602,0.4368,...,0.2061]] image_URL:'file:///.../DATSDP_1.jpg'

最后,将实体名称和对应的属性信息,转换成知识三元组格式,即属性三元组“<实体,属性名称,属性值>”的属性三元组格式进行存储。通过这种属性抽取方式,不仅可以高效地存储大量的实体属性信息,还能方便地进行扩展和更新。

## 4 知识图谱的关系抽取

关系抽取是指从非结构化文本中识别并抽取实体之间的语义关系,并通过知识三元组构建知识图谱中的关系边,丰富实体间的语义关联。本文针对基于振动信号的实体节点,通过智能诊断算法进行相关特征信息的故障识别,根据识别结果将特征信息与设备故障结果进行映射关联,进而实现实体之间的关系链接与知识结构化表达。

### 4.1 基于参数特征的故障识别

以 x、y 方向振动信号划分出的训练集与测试集为输入,分别提取出上述 8 个特征参数,得到 16 维的训练特征向量和测试特征向量;然后,将训练特征向量分别输入到 KNN、SVM、随机森林、贝叶斯分类器和决策树 5 种机器学习模型中进行训练;最后,将测试特征向量分别输入到对应的算法模型中进行故障类型的识别。每个诊断模型分别进行 5 次实验,提取出各个模型的准确率 Acc 与 KIA 值<sup>[17]</sup>绘制点线图,如图 5 所示。另外,计算 5 次测试结果准确率 Acc 和 KIA

值的平均值,用来评价各模型的诊断效率,5 种诊断模型的综合实验结果如表 11 所示。

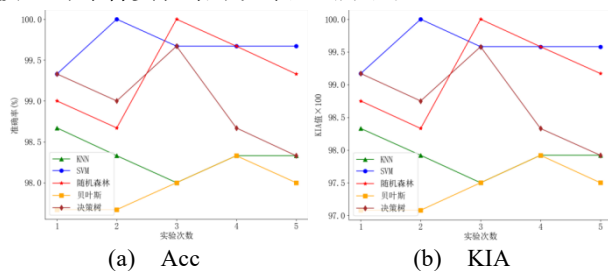


图 6 5 种诊断模型 5 次实验结果

Fig.6 Five experimental results of five diagnostic models

表 11 5 种诊断模型实验的平均结果

Table 11 Average results of five diagnostic model experiments

诊断模型	Acc(%)	KIA
KNN	0.9833	0.9791
SVM	0.9967	0.9959
随机森林	0.9933	0.9917
贝叶斯分类器	0.9793	0.9742
决策树	0.99	0.9875

对比分析可知,贝叶斯分类器模型的平均 Acc 为 0.9793,平均 KIA 值为 0.9742,在 4 种模型中效果是最差的。随机森林和决策树两种模型的准确率和 KIA 值均在 0.99 左右,但是从点线图中可以看到 5 次实验的 Acc 和 KIA 值并不稳定,波动幅度很大。其中 SVM 的故障分类效果最优,5 次实验准确率都在 0.90 以上,平均 Acc 达到 0.9967,KIA 值达到了 0.9959。

### 4.2 基于图像特征的故障识别

为验证 DATSDP 融合图像的实用性和可靠性,同时采用不同的图像变换方法和卷积网络模型进行对比实验。利用均值处理的融合方法得到融合信号后,分别利用灰度图 (grey)、马尔可夫转移场 (MTF) 和原始 SDP 方法三种图像生成方法,得到对应的融合图像数据集,其中 SDP 公式中的  $\theta$  分别设为  $90^\circ$ 、 $180^\circ$ 、 $270^\circ$  和  $360^\circ$ ,  $\xi$  设为  $40^\circ$ ,  $l$  设为 5;将 DATSDP 融合图像等四种图像训练集分别输入到 AlexNet、VGG 和 ResNet 三种经典卷积网络中进行训练后,将测试集输入到分别输入到对应的模型中,得到最终的分类结果,如图 6 所示。

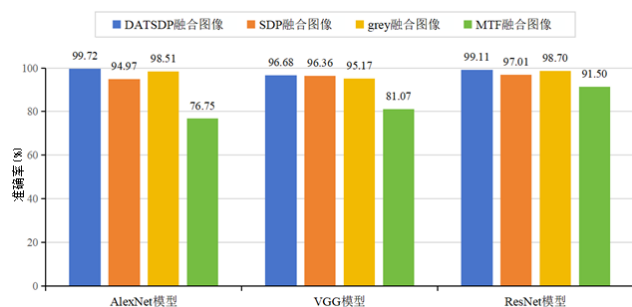


图7 不同图像特征在三种模型下的识别准确率

Fig.7 Recognition accuracy of different image features under three models

在四种图像中,MTF融合图像的准确率最低,尤其是在 AlexNet 模型中仅为 76.75%。另外, grey 融合图像和 SDP 融合图像的表现较为接近,在不同模型中的准确率波动较小。相比之下, DATSDP 融合图像在这三种模型中均取得了最高的准确率,分别是 99.72%、96.68%和 99.11%。这表明 DATSDP 方法在图像处理领域的表现,明显优于传统或其他图像提取技术,同时在处理多源数据时更具优势。

综上所述,利用最优类距比对多维度特征进行降维筛选,得到的参数特征在 SVM 分类器中识别效果最佳,准确率可达 99.67%。同时,利用均值处理对多方向信号进行融合处理,并通过 DATSDP 变换得到的融合图像,在 AlexNet 模型中的分类准确率达到到了 99.72%,展现出了更为稳定的诊断效率与性能表现。因此,利用 SVM 分类器和 AlexNet 模型分别对参数特征与图像特征进行故障识别,根据分类结果将相关实体节点进行关系链接,实现知识图谱的关系抽取。

### 4.3 关系抽取

根据上述参数与图像的故障识别结果,进行实体节点之间的关系定义与链接。首先,根据六级、五级和四级实体节点中的参数特征和图像信息,建立出实体节点之间的隶属关系;之后,将参数特征与 DATSDP 融合图像分别输入到 SVM 分类器和 AlexNet 网络中进行故障识别,根据对应的识别结果在四级与三级实体节点之间建立“参数预测结果”关系和“图像预测结果”关系;最后,在三级、二级和一级实体节点之间同样建立隶属关系,形成特征关系链,集合所有的关系链得到完整的知识图谱,如图 7 所示为特征关系链示意图。

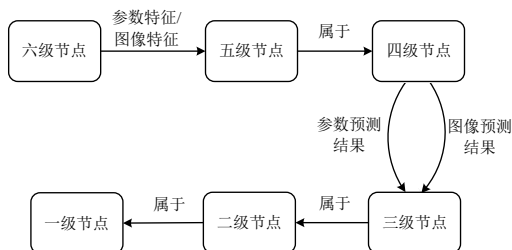


图8 特征关系链以及部分具体关系链示意图

Fig.8 schematic diagram of characteristic relationship chain and some specific relationship chains.

接着,进行渣浆泵数据集的实体及属性抽取、实体之间关系抽取,提取“<实体 1, 关系, 实体 2>”格式的知识三元组,并转化为表格文件进行存储,作为后续图谱创建的输入。

## 5 知识图谱的构建与检索

Neo4j 图数据库是一种强大的可扩展性数据库,在关系处理和实时数据更新方面具有出色表现。所创建的知识图谱,需要经常对图谱中的知识进行检索与更新,为方便有效的存储与查询,选择 Neo4j 图数据库作为搭建及展示知识图谱的载体。Cypher 查询语言是一种声明式的图形查询语言,专门设计用于 Neo4j 图数据库,用于直观地创建、读取、更新和删除节点、属性及关系。

对于包含属性描述的实体构建,例如属性三元组“<x 标准差, value, 0.2118777791171805>”,需要转换成“CREATE (m: x 标准差 {name: "x 标准差", value: 0.2118777791171805}) RETURN m” Cypher 代码,来生成对应的实体节点,并在 Neo4j 中进行展示,如图 8 所示。参考上述 Cypher 语言对六个层级的实体节点分别进行构建,得到包含丰富属性信息的实体节点。

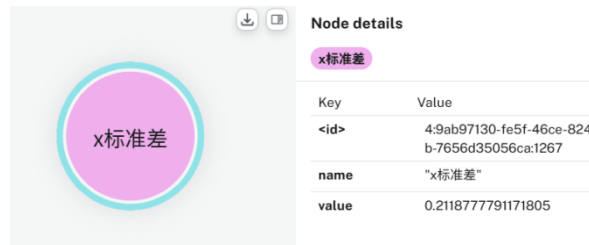


图9 实体节点的构建

Fig.9 Construction of Entity Nodes

其次,同样利用 Cypher 语言进行实体之间的关系创建,对于关系三元组“<x 标准差, 参数特征, x 方向信号>”,则需要转换成“MATCH (m: x 标准差 {value: 0.2231411065948729}), (n: x 方向信号 {feature\_value: [1239.1831482901132,0.2231411065948729,1440.6985234779556,734.8723430265726]}) CREATE (m)-[r: 参数特征]->(n) RETURN m, r, n”代码,利用 MATCH





关系三元组, 进而完成关系抽取。最终, 将三元组转换成 Cypher 语句并导入 Neo4j 图数据库, 构建出包含 36 010 个实体节点和 39 609 条实体之间关系的知识图谱, 为故障知识提供了结构化存储与高效检索。该方法不仅提升了诊断精度, 还为设备状态的可解释性分析提供了新范式。

未来研究将扩展该方法至多类型设备及复合故障场景, 并探索动态知识图谱的实时更新机制, 进一步提升工业诊断系统的智能化水平。

### 参考文献:

- [1] 何清波, 李天奇, 彭志科. 旋转机械故障诊断中的振动信号模型综述[J]. 振动.测试与诊断, 2024, 44(04): 629-639+821.  
HE Q B, LI T Q, PENG Z K. Overview of vibration signal model in fault diagnosis of rotating machinery[J]. Vibration. Test and Diagnosis, 2024, 44(04): 629-639+821
- [2] 周海军, 王超伟, 周国敬, 等. 基于随机森林的离心泵滚动轴承故障诊断[J]. 中国舰船研究, 2020, 15(03): 129-135.  
ZHOU H J, WANG C W, ZHOU G J, et al. Fault diagnosis of antifriction bearing of centrifugal pump based on random forest[J]. China Ship Research, 2020, 15(03): 129-135.
- [3] WANG S H, XIANG J W, ZHONG Y T, et al. A data indicator-based deep belief networks to detect multiple faults in axial piston pumps - ScienceDirect[J]. Mechanical Systems and Signal Processing, 2018, 112: 154-170.
- [4] 胡泽, 王晓杰, 张智博, 等. 基于 Hilbert-Huang 变换的离心泵故障诊断[J]. 电动工具, 2020, (02): 15-20.  
HU Z, WANG X J, ZHANG Z B, et al. Centrifugal Pump Fault Diagnosis Based on Hilbert-Huang Transform[J]. Electric Tools, 2020, (02): 15-20.
- [5] 庞新宇, 仝钰, 魏子涵. 一种 GAF-CNN 行星齿轮箱故障诊断方法[J]. 北京理工大学学报, 2020, 40(11): 1161-1167.  
PANG X Y, TONG Y, WEI Z H. A Fault Diagnosis Method for Planetary Gearbox Base on GAF-CNN [J]. Journal of Beijing Institute of Technology, 2020, 40(11): 1161-1167.
- [6] 侯玉叶, 闫海涛. 离心泵故障诊断的对称极坐标变换和图像形状特征方法[J]. 噪声与振动控制, 2023, 43(01): 116-121+172.  
HOU Y Y, YAN H T. Symmetric Polar Coordinate Transform and Image Shape Feature Method for Fault Diagnosis of Centrifugal Pumps[J]. Noise and Vibration Control, 2023, 43(01): 116-121+172.
- [7] SINGHAL. Official Google Blog: Introducing the Knowledge Graph: things, not strings[J]. Official Google Blog, 2012, 5(16): 3.
- [8] 冯钧, 杭婷婷, 陈菊, 等. 领域知识图谱研究进展及其在水利领域的应用[J]. 河海大学学报(自然科学版), 2021, 49(01): 26-34.
- FENG J, HANG T T, CHEN J, et al. Research status of domain knowledge graph and its application in water conservancy[J]. Journal of Hohai University (Natural Science Edition), 2021, 49(01): 26-34.
- [9] YOU B, YAN Y S, SUN Y G, et al. Method of Information Content Evaluating Semantic Similarity on HowNet[J]. Jisuanji Xitong Yingyong - Computer Systems and Applications, 2013, 22(01): 129-133.
- [10] NIU X, SUN X R, WANG H F, et al. Zhishi.me - Weaving Chinese Linking Open Data[C]. The Semantic Web-ISWC 2011: 10th International Semantic Web Conference, Bonn, Germany, October 23-27, 2011, Proceedings, Part II 10. Springer Berlin Heidelberg, 2011: 205-220.
- [11] 付雷杰, 曹岩, 白瑀, 等. 国内垂直领域知识图谱发展现状与展望[J]. 计算机应用研究, 2021, 38(11): 3201-3214.  
FU L J, CAO Y, BAI J, et al. Development status and prospect of vertical domain knowledge graph in China[J]. Computer Application Research, 2021, 38(11): 3201-3214.
- [12] HUANG J, CHEN Y H, LI Y, et al. Medical Knowledge-Based Network for Patient-oriented Visual Question Answering[J]. Information Processing & Management, 2023, 60(02): 103241.
- [13] JIANG L, LIU K P, WANG Y B, et al. Reinforced Explainable Knowledge Concept Recommendation in MOOCs[J]. ACM Transactions on Intelligent Systems and Technology, 2023, 14(03): 1-20.
- [14] HUANG X K, YANG G J, ZHANG Y Y, et al. Ontology Guided Multi-level Knowledge Graph Construction and Its Applications in Blast Furnace Ironmaking Process[J]. Advanced Engineering Informatics, 2024, 62: 102927.
- [15] 唐荻音, 丁奕州, 王轩, 等. 面向多源异构数据的航天器故障知识图谱构建方法[J]. 空间控制技术与应用, 2023, 49(04): 40-49.  
TANG D Y, DING Y Z, WANG X, et al. Fault knowledge graph construction method for spacecraft based on multi-source heterogeneous data[J]. Space Control Technology and Application, 2023, 49(04): 40-49.
- [16] 陈宏, 陈新财, 巩晓赞, 等. 基于知识图谱的风电机组诊断系统构建与应用[J]. 郑州大学学报(工学版), 2023, 44(06): 54-60.  
CHEN H, CHEN X C, GONG X Y, et al. Construction and application of wind turbine diagnosis system based on knowledge graph[J]. Journal of Zhengzhou University (Engineering Edition), 2023, 44(06): 54-60.
- [17] 刘晶, 高立超, 孙跃华, 等. 基于知识和数据融合驱动的设备故障诊断方法[J]. 郑州大学学报(理学版), 2022, 54(02): 39-46.  
LIU J, GAO L C, SUN Y H, et al. Fault Diagnosis Method for Equipment Driven by Knowledge and Data Fusion [J]. Journal of Zhengzhou University (Science Edition), 2022, 54(02): 39-46.
- [18] OKUZONO T, SHIMIZU N, SAKAGAMI K. Predicting absorption characteristics of single-leaf permeable membrane absorbers using finite element method in a time domain[J]. Applied Acoustics, 2019, 151: 172-182.
- [19] LIU Y, KANG H, GUO Z X, et al. A CNN-Based Ground Motion Selection Method Considering the Impact of Frequency and Time Characteristics of Ground

- Motions[J]. Journal of Structural Engineering. 2025, 151(4): 04025026.
- [20] HUANG K T , CHANG S H , LIN Y J. Microstructure, material characteristics and mechanical properties of CoMoNiTa3 medium-entropy alloys through the vacuum hot pressing process[J]. Powder Metallurgy, 2024, 67(2-3): 154-168.
- [21] 胡浩. 基于对称极坐标图像的旋转机械故障诊断方法研究[D]. 北京: 北京化工大学, 2022.
- HU H. Research on Fault Diagnosis Method of Rotating Machinery Based on Symmetric Polar Images[D]. Beijing: Beijing University of Chemical Technology, 2022.
- [22] 侯玉叶, 闫海涛. 离心泵故障诊断的对称极坐标变换和图像形状特征方法[J]. 噪声与振动控制, 2023, 43(01): 116-121+172.
- HOU Y Y, YAN H T. Symmetric Polar Coordinate Transform and Image ShapeFeature Method for Fault Diagnosis of Centrifugal Pumps[J]. Noise and vibration control, 2023, 43(01): 116-121+172.

1. 联系人: 秦志英;

2. 通讯地址(邮政编码): 050018;

3. 电子信箱、电话: qinzhy76@163.com 13623314063。