

# 基于人机共融的包装机故障诊断系统的设计与实现

张明琰<sup>1</sup>, 宋震<sup>2</sup>, 方世杰<sup>2</sup>

(1. 河南中烟黄金叶生产制造中心, 郑州 450000; 2. 杭州首域万物互联科技有限公司, 杭州 311199)

**摘要:** 随着工业 4.0 和物联网时代的来临, 基于经验和手册的设备维修方式已不能满足复杂设备维修的要求; 而传统的设备诊断系统往往只注重从物理传感器采集数据, 缺少人工经验的介入, 难以拥有自学习能力; 以 ZB45 烟草包装机为例, 提出一种具有自学习能力的人机共融新型故障诊断系统; 系统采用贝叶斯网络, 实现对传感器数据的自动推理; 通过自然语言处理模块与用户交互, 学习用户的维修经验, 并用来改进诊断效果; 提出了基于凸优化的标签选择方法, 根据观察到的现象推荐合适的标签, 以快速确定最可能的故障, 实现快速找到报警号码对应的故障源; 生产现场的实测数据表明, 所实现的系统可以有效降低万箱故障次数, 有效提高故障诊断精度, 降低故障排查时间。

**关键词:** 人机共融; 贝叶斯网络; 故障诊断; 凸优化; 标签选择

## Design and Implementation of Packing Machine Fault Diagnosis System Based on Human Machine Integrated Approach

ZHANG Mingyan<sup>1</sup>, SONG Zhen<sup>2</sup>, FANG Shijie<sup>2</sup>

(1. Henan Zhongyan Gold Leaf Production and Manufacturing Center, Zhengzhou 450000, China;

2. Hangzhou First State IoT Technology, Hangzhou 311199, China)

**Abstract:** With the advent of Industry 4.0 and the Internet of Things era, the methods of equipment maintenance based on experience and manuals can no longer meet the requirements of complex equipment maintenance. However, the systems of traditional equipment diagnosis often only focus on collecting data from physical sensors, lacking the experience of introducing the human factors, and it is difficult to have self-learning capabilities. Taking ZB45 tobacco packaging machine as an example, a new man-machine integrated fault diagnosis system with self-learning ability is proposed. The system uses Bayesian network to realize the automatic reasoning from the sensor data. Through the natural language processing module to interact with the user, learn the user's maintenance experience, and it is used to improve the diagnosis effect. A label selection method based on convex optimization is proposed, and the suitable labels are recommended according to the observed phenomena to quickly determine the most probable faults, and to quickly find the fault source corresponding to the alarm number. The measured data on the producing spot shows that the system can effectively reduce the number of faults for every 10 000 boxes, can effectively improve the accuracy of fault diagnosis and reduce the troubleshooting time.

**Keywords:** man-machine integration; Bayesian network; fault diagnosis; convex optimization; label selection

## 0 引言

常见的设备故障报警方法没有故障诊断的功能: 当传感器测到某物理量的值出现过大大偏差时, 系统会显示报警编号。用户必须对照手册, 自行诊断导致故障的原因。随着工业 4.0 和物联网时代的来临, 这种基于经验和手册的设备维修方式仅对于简单的设备有效, 但不能满足对复杂设备的要求。本文以烟草行业的 ZB45 型卷包机为研究对象, 基于故障树和贝叶斯推理网络, 旨在提出一种具有不断自我学习能力的人机共融的新型故障诊断系统。

本文研究方法的关键思想是要结合人的因素, 进行人机共融 (Human Machine Integration)<sup>[1]</sup> 的故障诊断, 并在

三方面尝试创新的技术方案。首先, 采用自然语言处理的方法, 通过历史工单和维修手册, 自动生成故障诊断用的贝叶斯网络和故障搜索用的知识图谱。对数千种故障的复杂系统而言, 这样不但可以大大节省人力, 也可以减少人为错误的可能, 并且可以快速处理新工单, 融入知识库。其次, 基于贝叶斯网络, 实现自学习功能。当工人确定故障源后, 可以点击“故障确认”按钮, 通知算法正确的结果。这样, 算法可以不断改进诊断精度。最后, 在故障处理中引入“人本传感器”的新概念, 让工人点击界面上的标签, 输入工人所观察到的现象。有效克服传感器数量远远小于故障数量的问题。让用户作为传感器, 帮助算法尽

收稿日期: 2021-11-26; 修回日期: 2021-12-22。

作者简介: 张明琰(1972-), 男, 河南新郑人, 大学本科, 高级工程师, 主要从事电器设备维修管理方向的研究。

引用格式: 张明琰, 宋震, 方世杰. 基于人机共融的包装机故障诊断系统的设计与实现[J]. 计算机测量与控制, 2022, 30(6): 23-31.

量唯一地确定故障源, 加速排错过程。本文提出的人机共融的诊断系统, 不但可以在人为工单数据中挖掘信息, 还可以通过在日常操作中学习工人经验, 同时借助人本传感器提高诊断效率。通过人机紧密合作, 实现了在工业 4.0 场景下, 新型故障诊断系统的范式。

## 1 ZB45 型包装机故障诊断研究现状

### 1.1 ZB45 型包装机及其故障诊断的必要性

烟草行业的 ZB45 型包装机是用于烟支卷接和包装的设备, 具有结构复杂、设备贵重、小故障停机次数多、连锁质量波动大、次生浪费较多的特点。其内部有约 10 000 个零部件, 约 2 000 种故障, 约 1 500 个 OPC 故障报警代码 (OPC 代码)。故障和 OPC 代码之间是多对多的对应关系。有约 1 300 种可导致停机的故障没有对应的 OPC 故障报警代码。ZB45 包装机在高速生产运行过程中, 包装机极易因各种原因产生短时高频停机, 根据对生产环境的统计, 单班停机次数平均约为 45 次, 会造成严重的生产效率损失。同时, 包装机的频繁启停会使多个关键工序存在加热、升温等不断调整的状态, 使包装工艺的质量稳定性变差。同时包装机在每次重启后设备都需自动剔出固定数量的烟支, 造成原、辅料浪费, 产生巨大的成本损失。

### 1.2 传统故障诊断方法存在的问题

包装机产生故障的成因很多, 往往会在任意时刻、任意部位产生不同类型的故障。以 ZB45 包装机的铝箔纸系统为例, 可能产生的故障就有数十种, 仅其中“铝箔纸撕不开”这一种故障, 又会有多种因素与其相关。由此可见, 对包装机的准确故障诊断难度极大。

在现阶段, 包装机故障主要依靠维修人员的经验, 对故障进行经验性的诊断, 主要技术手段是以手工查找纸版手册和维修记录为主, 故障诊断存在耗时长、具有主观性的问题, 极大依赖于人员经验, 技术传承性较差, 往往会导致大量故障停机时间。特别是对于一个 OPC 代码对应多个可能故障的情况。而维修手册只记录各种故障可以导致的现象, 并且手册有几千页之多。不但查找困难, 而且很多故障诊断的经验没有书面记录。有经验的工人要根据故障代码, 结合自己的观察, 及历史故障频次来判断最有可能的故障。而这种被高频“依赖”的经验常常没有书面记录, 动辄需要五年、十年的经验积累。

随着制造业信息化进程的推进, 卷包机数据采集系统实现了从 ZB45 包装机提取大量实时运行数据, 包括运行参数、故障统计、台时产量、物料消耗等信息, 同时也建立了基于 EAM 系统相关的设备故障维修记录。但是两个系统独立运行, 系统与系统间缺乏有效的关联策略来实现数据和信息的融合, 需要对获取的数据进行更深层次的处理和分析, 从各系统数据进行关联挖掘, 以得出有效、可固化传承的故障诊断及设备维护类知识, 并应用到生产过程中。

### 1.3 相关技术背景

#### 1.3.1 常见的故障诊断方法

传统的复杂设备故障的诊断主要包括如下方法:

1) 故障树: 通过故障树的方法<sup>[2-3]</sup>建立从传感器可测的故障现象到故障的自动推理引擎。这种方法也常被称做专家系统。如图 1 所示的故障树, 不同的故障可以通过逻辑“与”、“或”、“非”的方式关联, 共同决定下一级故障发生与否。故障树方法可以用于系统可靠性分析, 及单一故障情况下的故障诊断。

2) 贝叶斯网络: 另一类方法是通过如贝叶斯网络<sup>[2,4]</sup>的统计学方法, 对多个故障及其故障概率进行自动推理。这类方法需要用户构建以故障和现象为节点的有向无环图 (DAG, directed acyclic graph), 并输入条件概率表 (CPT, conditional probability table) 和先验概率以表征不同节点之间的连接关系。如果有足够历史数据, 条件概率表和先验概率都可以从数据中得出。这种方法比故障树的表现力更强。故障树可以转换为贝叶斯网络的形式<sup>[2]</sup>, 但反之则不一定成立。贝叶斯网络的方法可用于大型生产制造系统<sup>[4]</sup>、楼宇空调<sup>[5]</sup>的故障诊断及故障冗余控制系统<sup>[6]</sup>。基本的朴素贝叶斯网络假定故障之间独立出现, 即故障出现的概率不会相互影响, 同时, 故障出现的概率不随时间而变化。动态贝叶斯网络扩展了朴素贝叶斯网络的时间域上的表现力, 从而允许出现时变的概率分布。

3) 观测器: 如果要通过观测模拟量来分析故障情况, 可以采用观测器的方法, 并通过在线传感器数据估计系统内部状态。例如: 卡尔曼滤波器 (Kalman filter) 或粒子群滤波器 (particle filter) 是常见的基于模型的系统状态观测器。对于只有模型结构, 但不了解参数的系统, 还可以采用在线模式识别的方法, 例如在线 ARIMA 或 N4SID 的算法<sup>[7]</sup>。

4) 数字孪生: 近年来, 数字孪生 (digital twin)<sup>[8-10]</sup>的概念得到研究者和工业界的重视, 并从 3D 模型延展到在线监测应用。对于大型的工业控制系统, 如石化企业的分布控制系统 (DCS, distributed control system)<sup>[10]</sup>及暖通空调系统<sup>[11]</sup>, 想要手动建立系统的动力学模型是非常困难的。对这类应用, 通过组合子系统的数字孪生而构建全系统的动力学模型是个更好的方案。文献 [12] 提出了采用在线数字孪生的方法, 对暖通空调设备进行在线故障监督和诊断。

#### 1.3.2 自然语言处理技术和知识图谱

近年来, 以深度学习为基础的自然语言处理得到了广泛的工业应用。对设备故障诊断而言, 设备维修工单记录和维修手册是故障诊断系统建模和分析的重要数据来源, 其描述形式基本上是自然语言。然而, 直接通过阅读自然语言, 手工建模的方法成本高, 时间长。利用自然语言处理技术是提高故障诊断系统建模的有效手段。早期的应用主要以针对关键词的统计为主<sup>[13]</sup>。美国海军采用了自然语言处理的方法对 V-22 鱼鹰飞机的维护资料进行了系统的分析整理<sup>[14]</sup>。通过基于 Python 的 NLTK 自然语言处理工具包<sup>[15]</sup>进行分词, 该方法识别了维修记录中的自然语言记号, 并对核心词进行了统计学分析。在此基础上, 可以对历史上相似的故障维护工单进行聚类, 并通过统计方法分析改进维修的手段。针对烟草机械维护的自然语言处理应用还



图 1 故障树部分展示

很少见。文献 [16] 讨论了烟草机械维护数据库的自然语言处理方法, 采用常规的 TF-IDF 及 TextRank 模型进行数据建模, 然后通过权重矩阵比较实际故障记录与故障数据库中已有故障记录的相似程度, 由此进行对同类故障记录的模糊查找, 但是其本身并不直接用于故障诊断。当涉及到具体行业知识时, 这种模糊查找方法效果不太理想。

知识图谱<sup>[17]</sup>是一种描述实体之间语义关系的有向图结构。传统的知识图谱一般存于图数据库中, 如 Neo4J, Dgraph 等。近年来, 随着人工智能和深度学习的兴起, 出现了神经网络和知识图谱相结合的流派。例如, DeepWalk<sup>[18]</sup>可以通过深度学习来学习网络的结构。通过深度图卷积网络 (DGCN, deep graph convolution network) 学习轴承的振动信号, 文献 [19] 提出了一种可以进行故障原因诊断的学习系统。开源软件 Deep Graph Library (DGL)<sup>[20]</sup>, 可以把多种网络结构的数据向量化, 并嵌入到神经网络中。利用已有的知识图谱构建技术。我们通过维护手册和工单数据构建知识图谱, 并由此生成贝叶斯网络模型, 用于故障诊断。

## 2 人机共融的故障诊断系统

### 2.1 人机共融的新型交互方式

人机共融被认为是工业 4.0 时代的重要场景之一<sup>[1]</sup>。关于工业机器人的人机共融<sup>[21]</sup>及离散生产线的人机共融<sup>[22]</sup>已

经有一些成功经验。但对于故障诊断系统的人机共融方式还未见相关报道。我们以 ZB45 卷包机为实验对象, 提出一种故障诊断系统, 旨在提高故障诊断效率、降低故障停机时间, 并对故障的原因、部位与维护策略进行全面分析, 进而实现增效、稳质和降本的效果。其核心思想是通过人机共融的有机合作引导工人监督系统不断学习, 解决现有故障建模中缺乏文字描述经验的难点。本系统人机共融的特点主要体现在: 1) 建立基于贝叶斯网络的可扩展故障诊断架构, 可以利用 OPC 代码、用户标签选择、传感器信号等多模态信息, 进行自动故障诊断推理; 2) 通过人机共融的有机合作, 向用户提供标签, 并接收反馈; 学习用户潜在经验, 不断提高诊断精度; 3) 通过自然语言处理的方法, 分析历史工单数据, 自动构建推理用的贝叶斯网络, 减少手工建模时间。

系统的整体维护工作流程如图 2 所示, 其中, 机械设备的传感器采集信号处理后产生各类报警信号, 并利用相关事件检测算法, 触发事件后传送给贝叶斯网络。事件监测算法可以看作与内置的 OPC 代码报警相同, 都是触发事件的布尔量。而维修人员产生历史工单, 经过自然语言处理模块进行分析, 以生成故障知识图谱, 并用于初始化贝叶斯网络。工单的信息经过自然语言处理, 生成初选标签, 经凸优化过滤后, 产生足够的精选标签, 供维修人员选择。

标签中一般包括维修人员可以看到或听到的各种现象。作为对传感器信号的有效补充，标签可以看成另一种“人工”的传感器。它们也可以触发贝叶斯网络的事件。

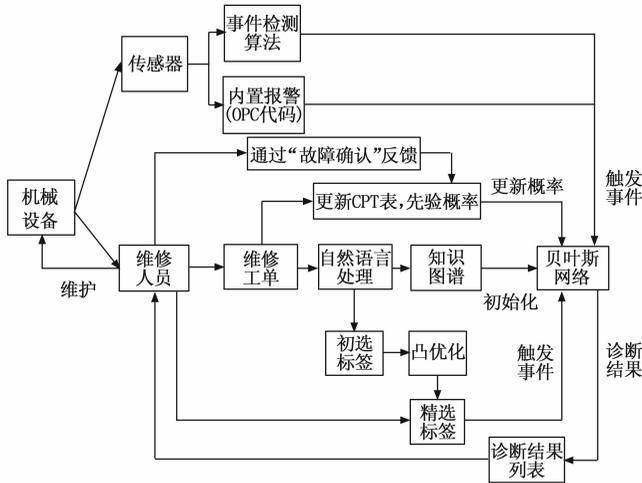


图 2 人机共融的卷包机故障诊断系统流程图

下面将依次介绍诊断系统中核心的贝叶斯网络推理运算、知识图谱应用和标签选择的凸优化算法，然后给出诊断系统的整体界面功能。

### 2.2 故障树建模

目前在 ZB45 包装机应用中积累了维修工人师傅的大量维修记录，但是记录存在标准不一、记录不全、故障重复等问题。在维修工人师傅经验的帮助下，从数十万的维修数据中，清洗整理出 1 751 条完整的故障维修记录，并将故障现象和机器位置相关联，形成故障树，同时对每个故障所对应的故障编码进行对应，通过这种关联，实现对故障的推送。故障树具体结构参见图 1，故障树与故障编码的关联如图 3 所示。

### 2.3 基于贝叶斯网络的推理运算

因为 OPC 代码和故障之间是复杂的多对多关系，难以用简单的逻辑表达，所以本文采用贝叶斯网络进行自动推理。首先，通过图 3 说明其推理的机理。贝叶斯网络的概念在图 4 (a) 和 (b) 中显示。其中，贝叶斯网络图 4 (a) 表示了故



图 3 故障树与故障编码关联

障和故障现象的因果关系; 而图 4 (b) 是条件概率表 (CPT, conditional probability table)。CPT 表是对贝叶斯网络图中箭头的定量注释。如, 从事件 A 到事件 C 的箭头第一行表示当故障 A、B 出现的概率为 0% 时, 现象 C 不出现的概率为 100%, 而 C 出现的概率显然为 0%。贝叶斯网络图中的每个节点是个事件, 如故障或现象。图中箭头表示从原因到结果的相关关系。本文中, 故障是指不能通过传感器数据直接判断的源头故障, 而现象是指可以通过传感器数据直接测量的现象。对诊断系统而言, 主要注意的是多个故障导致一个现象的网络, 对这类网络的诊断显然需要一定经验。如图 4 (a) 所示, 如果只知道 C 发生了, 显然无法唯一地确定究竟是故障 A, 还是故障 B 发生了。但是算法可以利用贝叶斯定理分别估计两种故障发生的概率。

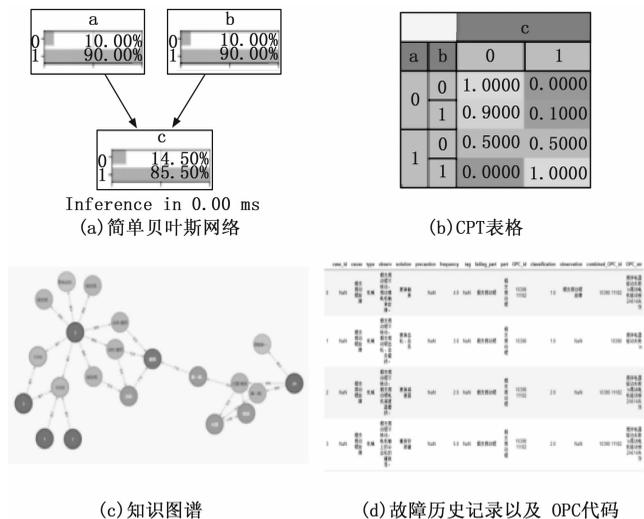


图 4 贝叶斯网络

假定故障 A 和故障 B 都可能引入事件 C。其中事件是可检测的, 而故障是不能直接测量的。当事件 C 发生后, 我们可以根据贝叶斯公式估计 A 出现故障的概率如公式 (1):

$$P(A | C) = \frac{P(C | A)P(A)}{P(C | A)P(A) + P(C | B)P(B)} \quad (1)$$

其中:  $P(A), P(B)$  是故障的先验概率, 可以通过统计历史上 A 和 B 发生的频率计算;  $P(C | A), P(C | B)$  分别是 A 或 B 发生时 C 发生的条件概率。如果根据行业知识, 当 A 发生时 C 一定出现, 则  $P(C | A)$  为 100%。当缺少行业知识, 或者希望通过大数据提高故障诊断的准确性时, 可以从维护日志 (工单) 数据中生成 CPT 表, 进而计算条件概率。统计一段时间内各事件发生的次数后, 则条件概率  $P(C | A)$  可以计算如公式 (2):

$$P(C | A) = \frac{N_{AC}}{N_A} \quad (2)$$

其中:  $N_{AC}$  是一段时期内事件 A、C 同时发生的次数, 而  $N_A$  是事件 A 发生的次数。类似的, 可以计算 CPT 表内其它条件概率, 如  $P(C | B)$  可以计算如公式 (3):

$$P(C | B) = \frac{N_{BC}}{N_B} \quad (3)$$

而  $N_{BC}$  是事件 B、C 同时发生的次数, 而  $N_B$  是事件 B 发生的次数。先验概率  $P(A), P(B)$  的计算方法分别为公式 (4)、公式 (5):

$$P(A) = \frac{N_A}{N} \quad (4)$$

$$P(B) = \frac{N_B}{N} \quad (5)$$

其中:  $N_A, N_B, N$  分别为事件 A、B 和所有事件发生的次数。

上述的贝叶斯网络有数千个节点, 手动构建网络工作量大。我们引入知识图谱来自动构建大型贝叶斯网络。知识图谱的构建需要获知实体和实体之间的关系, 即所谓的三元组信息。获取三元组有两种方式: 1) 从设备维修手册中获取结构化的三元组关系; 2) 从工单记录数据库中描述故障原因的句子中提取三元组。第二种方法要用到自然语言处理中的依存句法分析。首先对句子做分词和词性标注, 然后分析词与词之前的关系得到句法分析树, 然后遍历句法分析树, 得到可能的三元组。综上, 可以从数据中得到了若干三元组, 比如: (故障, 发生, 现象)、(部位, 属于, 部位) 等, 其中, “故障”、“现象” 等名词作为实体, “发生”、“属于” 等动词属于关系。实体作为拓扑图中的节点, 而关系作为拓扑图中的有向边, 构建了一张有向图, 即我们的知识图谱。之后, 通过 Python 程序构建知识图谱。部分知识图谱呈现于图 4 (c)。

知识图谱中有关故障和现象的三元组可以构建贝叶斯网络的结构。其它的三元组可以用于模糊搜索类功能。我们采用 pyArgum<sup>[23]</sup> 实现贝叶斯网络的构建和推理。软件包 pyArgum 的后端是由 C++ 实现, 速度快。而其丰富的可视化功能可以很好地解释推理的过程, 利于开发调试。

### 2.4 基于凸优化的标签选择算法

根据图 2 的流程图, 为了提高诊断效率, 采用中文分词工具 “结巴分词”<sup>[24]</sup>, 对如图 4 (d) 中的维修记录进行自然语言分析。图中 observation 和 OPC string 两列分别表示维修开展前工人经常观察到的现象及维修结束后工人对故障的总结。Observation 一列中可以由分词方法生成初选标签。结巴分词主要是基于统计方法来实现分词, 如隐马尔科夫模型 (HMM, hidden markov model)。利用结巴分词的用户词库功能, 我们建立了领域特有词库和停用词库, 用于识别行业专业词汇及可忽略的词汇。经过分词, 针对每个 OPC 代码得到如表 1 所示的初选标签。表中每行对应一个故障源, 每一行的标签组合都不同, 意味着用户只要在图 7 中选择足够的标签, 就可以唯一的确定故障源。但是如果太多的标签, 会影响用户的工作效率。因为故障总量数以千记, 显示需要自动化的算法。标签选择算法的目标是 “尽量减少图中总标签个数”, 同时 “保持每行有独立的标签组合”。表 1 是标签选择之后的效果。这些工单都对应同一 OPC 代码的 9 个不同故障。其中每行工单描述了一个不同的故障现象。初选标签是由自然语言处理选择的。而精选标签是通过基于凸优化的标签选择算法, 在初选标

签组中选择的。在初选标签组中共有 37 个独特标签，而精选标签组共有 11 个独特标签。

表 1 标签选择例子

原始工单	初选标签	精选标签
包装机 3Q1 跳闸, 有 24V 短路。	包装机 3Q1 跳闸 24V 短路	3Q1
CH 小包透明机上线路短路使空气开关 3Q1 跳闸。	CH 小包透明机上线路短路 空气开关 3Q1 跳闸	CH 线路 空气开关 3Q1
Q17 下端有短路线路使 Q17 跳闸。	Q17 下端 短路 线路 跳闸	Q17 线路
防护罩联锁开关磨损短路使 2Q17 跳闸。	防护罩 联锁 开关 磨损 短路 2Q17 跳闸	2Q17
Q22 下端出线有短路现象使 Q22 跳闸。	Q22 下端 出线 短路 现象 跳闸	Q22
因 CH 主电机故障, 引起电气柜内跳闸。	CH 主电机 故障 引起 电气 柜内 跳闸	CH
主机油泵电机 2M733 轴承损坏。	主机油泵 电机 2M733 轴承 损坏	2M733
主机油泵电机 2M733 过载, 引起 2Q733 跳闸。	主机油泵 电机 2M733 过载 引起 2Q733 跳闸	2M733 2Q733
空气开关 Q21 跳闸, 停机时涂胶器不转。	空气开关 Q21 跳闸 停机时 涂胶器 不转	空气开关 Q21

下面对标签选择算法进行数据建模。首先, 需要用 一个数学指标表示“保持每行有独立的标签组合”的思想。在此先通过比较简单的例子说明其建模过程, 再拓展概念。假设 有 3 个标签,  $X_1, X_2, X_3$ , 记  $X_1 = 1$  为选择此标签, 而  $X_1 = 0$  为不选此标签。显然, 所有标签选择的组合是如图 5 中单位立方体的每个顶点。顶点 A 的角度坐标为  $X_1 = 0, X_2 = 1, X_3 = 0$  即  $[0, 1, 0]$ 。说明当标签数  $N_t = 3$  时, 最多可以表达 8 个独立的故障源, 记为  $N_f = 2^N = 8$ 。引入向量  $a_i$  为第  $i$  个故障对应的标签组合, 矩阵  $A = [a_1, a_2, \dots]$ 。于是每个  $a_i$  对应图中一个顶点。如故障 1 对应  $a_1 = [1, 0, 0]^T$ , 即点 C。故障 2, 3 分别对应 A 和 H。对应单位立方体的体积:

$$A_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

当然, 这样定义的局限性在于  $A_1$  只能是方阵。为了解决这一问题, 我们尝试引入统计学理论的费舍尔信息矩阵 (FIM, fisher information matrix)<sup>[25-26]</sup>  $M$ 。当  $M$  为  $A_1 A_1^T$ , 费舍尔信息可理解为立方体体积的平方:  $|\det M| = |\det A_1 A_1^T| = 1$ 。当有两个故障的标签重复时, 例如:

$$A_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

这时  $|\det M| = |\det A_2 A_2^T| = 0$ , 表明  $M$  可以表征标签出现重复。这样, 可以分析任意多个故障组合。例如, 当  $M$  为所有顶点时,  $|\det M| = |\det A A^T| = 32$ 。

下面, 针对标签选择的应用, 重新定义  $M$  如下。

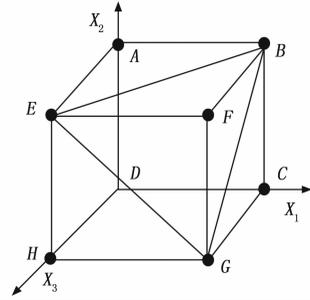


图 5 标签选择算法的物理意义

定义 1 (标签选择信息矩阵):

$$M = \sum_{i=1}^{N_f} p_i a_i a_i^T = A P A^T$$

$$P = \begin{bmatrix} p_1 & 0 & 0 & \dots \\ 0 & p_2 & 0 & \dots \\ 0 & 0 & p_3 & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix} \quad (6)$$

其中:  $a_i$  是第  $i$  个故障的标签选择列向量;  $A = [a_1, a_2, \dots]$ ;  $P_i$  是个布尔量代表第  $i$  个标签组合是否被选中。

为了保证线性变化后优化解不受影响, 在最优实验设计时一般用 D-最优 (D-Optimality) 做为目标函数, 记做  $\psi(M) = -\log \det(M)$ , 这样, 可以基于以前的传感器选择算法, 采用以下的凸优化算法<sup>[25]</sup>求解:

$$\min_{p_i} \psi(M) = -\log \det \left( \sum_{i=1}^{N_f} p_i a_i a_i^T \right)$$

$$s. t. \quad \sum_{i=1}^{N_f} p_i = 1$$

$$p_i \in [0, 1]$$

$$p_i \in R \quad (7)$$

其结果的输出为  $P_i$ , 其值接近于 0 或 1。其中  $N_f$  为不超过  $\frac{\log(N_f)}{\log 2}$  的最大整数。之后, 选取  $k$  个最大的  $P_i$ , 选择其对应的标签。从  $k = N_f$  开始。如果有解, 则令  $k = k - 1$ , 以此类推。直到没有可唯一选择的标签。

### 2.5 人机共融视域下的故障诊断及设备维护专家系统实现

基于上述原理, 本文开发了完整的人机共融故障诊断及设备维护专家系统。首先, 用户可以通过图 6 的首页大屏浏览车间情况。全车间的数据以图文并茂的方式分别在左、中、右三列展示出来。其中左列显示各线路设备运行情况, 中列显示设备作业效率, 右列展示故障历史数据。当设备出现新的故障时, 右下角推送信息可以提醒用户注意。当用户在左侧导航选择一个正常工作状态的设备时, 可以看到如图 6 的设备页大屏。在中间即时地显示了历史故障记录、OPC 代码、故障说明、故障时间、排查时间、停机时间、有效作业时间等信息。

当某设备出现 OPC 报警时, 用户可以点击左侧导航条上相应的设备图标, 并进入如图 7 所示的自动诊断页面。右侧显示了推荐的诊断结果, 包括置信度、故障信息、维



(a) 首页大屏



(b) 设备页大屏

图 6 可视化大屏

护方法、“故障确认”按钮。一个 OPC 代码可能对应多个故障。为了减少右侧诊断结果的数量,中间栏下方显示了故障标签。标签是通过算法在维修记录中里抽取的关键词。其中很多是用户可以直接观察到的现象。用户可以通过选择合适的标签,帮助系统更精准的诊断。当用户经过检修,确认正在发生的故障后,可以点击相应的“故障确认”按钮以通知系统。系统可以学习各故障出现的概率。这样可以帮助系统未来再推荐更准确的诊断结果。因为传感器数量有限,有时设备停机后没有 OPC 代码。这时,用户需要通过如图 7 的手动诊断界面来定位故障。工人可以通过图 7 (a) 的下拉菜单选择故障位置,之后用中间的标签缩小推荐范围,其过程和自动选择类似。



(a) 自动诊断界面



(b) 手动诊断界面

图 7 诊断界面

### 3 诊断系统的生产成效

通过对比系统部署前后月份的万箱故障数据,如表 2 和表 3 所示。从中可以看到,系统部署后一个月的故障次数为 46 次,万箱故障次数为  $46/69\ 296.4 \times 10\ 000 = 6.63$ ,而部署前一个月的故障次数是 64,万箱故障次数为  $64/66\ 232.2 \times 10\ 000 = 9.64$ ,同比万箱故障次数下降 31.2%。

表 2 系统部署后一个月的万箱故障数据

行号	常用名	制度时间	运行时间	故障次数	故障停机时间	故障率/%
1	27 号包装机(97 细)	544.5	463.27	1	1	0.18
2	17 号包装机	544.5	458.09	3	3.5	0.64
3	03 号包装机	544.5	464.74	5	5.5	1.01
4	16 号包装机	544.5	458.42	4	5	0.92
5	05 号包装机	544.5	459.46	4	1.66	0.30
6	19 号包装机	544.5	458.94	0	0	0.00
7	02 号包装机	544.5	457.53	2	1	0.18
8	15 号包装机	369	318.22	12	14.33	3.88
9	18 号包装机	544.5	452.72	3	3.67	0.67
10	04 号包装机	544.5	459.15	0	0	0.00
11	20 号包装机	544.5	455.75	5	17.17	3.15
12	23 号包装机	392.5	327.9	1	2	0.51
13	25 号包装机(97 细)	544.5	445.52	1	0.5	0.09
14	26 号包装机(97 细)	544.5	459.92	2	1.01	0.19
15	24 号包装机(97 细)	544.5	462.54	3	1.5	0.28
16	13 号包装机(97 细)	544.5	452.5	0	0	0.00

表 3 系统部署前一个月的万箱故障数据

行号	常用名	制度时间	运行时间	故障次数	故障停机时间	故障率/%
1	27 号包装机(97 细)	562	430.51	1	0.5	0.09
2	17 号包装机	104.5	68.75	2	0.66	0.63
3	03 号包装机	570.5	438.4	4	2.33	0.41
4	16 号包装机	570.5	434.31	5	3.7	0.65
5	05 号包装机	330.5	261.68	4	2.33	0.70
6	19 号包装机	450.5	340.69	4	9	2.00
7	02 号包装机	570.5	434.33	11	5.74	1.01
8	15 号包装机	522.5	399.65	7	9.3	1.78
9	18 号包装机	570.5	450.11	2	1	0.18
10	04 号包装机	457.5	354.8	6	7.07	1.55
11	20 号包装机	562	431.71	3	3	0.53
12	23 号包装机	562	425.8	3	1.33	0.24
13	25 号包装机(97 细)	562	424	0	0	0.00
14	26 号包装机(97 细)	562	421.95	9	22.66	4.03
15	24 号包装机(97 细)	562	441.81	1	1	0.18
16	13 号包装机(97 细)	332	219.77	2	2.33	0.70

通过分析该时间段内发生频次最多的前 20 个故障数据,得到部署后的故障维修时间平均降低了 22.22%。故障统计汇总数据如表 4 所示。

表 4 系统使用前与使用后故障处理情况对照表(频次前 20)

OPC 代码	使用前出现次数	使用前平均故障时间/s	使用后出现次数	使用后平均故障时间/s
10713	69	4 920.1	56	3 808.4
10729	41	4 891.9	64	4 891.1
10183	33	2 019.1	27	1 692.1
10136	32	2 029.1	15	1 751.8
10454	26	4 160.9	18	2 491.5
10727	15	4 314.7	18	2 700.8
10741	10	2 256	19	1 536
10739	10	6 546.5	15	5 091
10728	12	4 399.1	13	3 171.3

故障诊断系统的性能指标中,故障辨识的精确性是指对故障的大小机器事变特性估计的准确程度。我们通过统计故障确认位置在系统推荐的前三故障位比例,作为系统的故障辨识精确率,如式(8)所示。表 5 是 OPC 编号为 10136 的 CH 输入阻塞故障确认位置的详细数据。

$$\text{故障辨识精确度} = \frac{\text{该故障确认位置在推荐前三的次数}}{\text{该故障出现的总次数}} \quad (8)$$

如表 5 所示的故障辨识精确度为  $14 \div 15 = 93.33\%$ 。表 6 列出了总共 10 个故障代码及其对应的辨识精确度统计情况。

表 5 故障位置确认情况样例(CH 输入阻塞故障)

故障开始时间	故障确认位置	机器编号
2020-07-16 18:23	5	BZJZ_23
2020-07-18 4:32	2	BZJZ_16
2020-07-19 6:34	1	BZJZ_27
2020-07-22 6:40	1	BZJZ_17
2020-07-22 7:11	3	BZJZ_23
2020-07-25 6:45	1	BZJZ_16
2020-07-25 23:40	2	BZJZ_20
2020-07-26 17:37	2	BZJZ_13
2020-08-02 23:42	1	BZJZ_24
2020-08-15 7:48	1	BZJZ_27
2020-08-16 5:25	1	BZJZ_18
2020-08-16 5:30	1	BZJZ_23
2020-08-19 12:30	1	BZJZ_18
2020-08-31 7:18	2	BZJZ_15
2020-08-31 8:20	1	BZJZ_13

表 6 10 个故障代码的精确度统计结果

OPC 代码	中文故障名	精确度/%
10713	烟库无烟	94.20
10729	4 号轮无内衬纸	92.19
10183	CH 拉丝断或无	100
10718	无铝箔纸	86.96
10136	CH 输入阻塞	93.33
10454	烟道阻塞	100
10727	缺拉纸	88.89
10741	15 点无盒片	89.47
10739	盒片吸附引起的阻塞	86.67
10728	拉纸检测失效	92.21

## 4 结束语

面向复杂设备的故障诊断问题,本文以 ZB45 包装机故障诊断为考察目标,提出了一种实现人机共融的故障诊断系统。系统整合了自然语言处理技术、知识图谱技术和贝叶斯网络理论,主要创新点如下:1)通过对历史工单的自然语言处理,自动生成知识图谱和贝叶斯网络;2)针对故障代码(OPC 代码)少于故障源的问题,鼓励用户通过选择标签的方式来提高诊断精度。同时,利用用户反馈,提高贝叶斯网络诊断精度;3)结合自然语言处理和凸优化的方法,提供少而精的标签。通过对历史工单的自然语言处理,进行标签初选。随后采用凸优化算法选择恰好足够的标签给用户选择。通过系统实际运行测试,证明该系统能有效降低设备的万箱故障次数,有效减少故障维修时间,提高了设备故障诊断效率。

### 参考文献:

- [1] 於志文,郭 斌. 人机共融智能 [J]. 中国计算机学会通讯, 2017, 13 (12): 64-67.
- [2] 王广彦,马志军,胡起伟. 基于贝叶斯网络的故障树分析 [J]. 系统工程理论与实践, 2004, 24 (6): 78-83.
- [3] RUIJTERS E, STOELINGA M. Fault tree analysis: a survey of the state-of-the-art in modeling, analysis and tools [J]. Computer Science Review, 2015, 15: 29-62.
- [4] CARBERY C M, WOODS R, MARSHALL A H. A Bayesian network based learning system for modeling faults in large-scale manufacturing [C] //2018 IEEE International Conference on Industrial Technology (ICIT), 2018: 1357-1362.
- [5] XIAO F, ZHAO Y, WEN J, et al. Bayesian network based FDD strategy for variable air volume terminals [J]. Automation in construction, 2014, 41: 106-118.
- [6] VILLEZ K, VENKATASUBRAMANIAN V, NARASIMHAN S. Bayesian inference for fault-tolerant control [C] // International Symposium on Resilient Control Systems, IEEE, 2009: 51-53.
- [7] 欧阳霄,杨 辉,朱东济,等. 基于 ARIMA 和神经网络的单脉冲雷达设备故障预测研究 [J]. 计算机测量与控制, 2021, 29 (10): 83-87.
- [8] WANG J, YE L, GAO R X, et al. Digital twin for rotating machinery fault diagnosis in smart manufacturing [J]. International Journal of Production Research, 2019, 57 (12): 3920-3934.
- [9] XU Y, SUN Y, LIU X, et al. A digital-twin-assisted fault diagnosis using deep transfer learning [J]. IEEE Access, 2019, 7: 19990-19999.
- [10] 李 智,姜 悦,蔡斐华等. 数据驱动的飞行器智能故障诊断系统研究 [J]. 计算机测量与控制, 2021, 29 (1): 50-53.
- [11] LU Q, XIE X, PARLIKAD A K, et al. Digital twin-enabled anomaly detection for built asset monitoring in operation and maintenance [J]. Automation in Construction, 2020, 118:

