

ICS 35.240.50

CCS J 07



团体标准

T

T/CI 974—2025

制造业产品的全生命周期数据知识推理技术规范

Full life cycle data knowledge reasoning technical specifications for manufacturing products

2025 - 04 - 21 发布

2025 - 04 - 21 实施

中国国际科技促进会 发布

目 次

前言	II
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 知识获取，整合和表示	2
4.1 整体架构	2
4.2 数据接入层	2
4.3 数据清洗层	2
4.4 知识整合与表征层	2
4.5 知识推理层	3
4.6 应用层	3
5 算法构建要求	3
5.1 数据采集	3
5.2 数据清洗和预处理	3
5.3 整合不同来源的数据	3
5.4 知识表示以及知识图谱的构建	3
5.5 知识推理	4
5.6 评价指标	4
6 应用场景	5
6.1 故障溯源	5
6.2 供应链调控	6
6.3 决策辅助	6

前 言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由南方科技大学提出。

本文件由中国国际科技促进会归口。

本文件起草单位：南方科技大学、吉林大学、长春理工大学、中国第一汽车股份有限公司、北京大学深圳研究生院、东南大学、深圳市骏嘉科技发展有限公司、深圳市铠硕达科技有限公司、深圳浑沌数字化实验室科技有限公司、深圳树米网络科技有限公司、汕头大学、吉林省卡思特科技有限公司、吉林省中云数讯科技股份有限公司、联通（吉林）产业互联网有限公司、吉林省环境科学研究院。

本文件主要起草人：宋轩、来晓、冯德帆、谢洪彬、张家祺、赵家兴、雷裕珍、刘双雪、黄天律、王庆领、张静远、李佳奇、范子沛、袁飞、王泽恺、杨华民、张浩然、张凌宇、张昕、宋小龙、刘妍、周时莹、李长龙、高仕宁、孔祥明、孙宗姚、陈欣、陈瑶、王为、田博。

制造业产品的全生命周期数据知识推理技术规范

1 范围

本文件规定了针对标准制造业流程中全生命周期的数据进行推理的知识获取、整合和表示，算法构建，应用场景等要求。

本文件适用于标准实施后新产生的包括生产、加工、组装、售卖在内的制造业流程全生命周期数据所构成的知识图谱对于现有知识的纠错以及对于未直接包含的知识推理的流程，能作为标准制造业产品全生命周期数据的知识推理的技术参考，并给出了应用场景整体流程规范。

2 规范性引用文件

本文件没有规范性引用文件。

3 术语和定义

下列术语和定义适用于本文件。

3.1

实体 entity

在知识图谱中具有独特标识和语义含义的事物或概念。这些实体可以是现实世界中的个体、抽象概念、事件、地点、时间等各种元素，它们通过具体的属性和关系与其他实体相互联系和描述。

3.2

关系 relationship

用于连接实体之间关联、描述或者表达相互之间关系的概念，其说明两个或多个实体之间的联系、依赖、关联或者其他形式的连接方式。这些关系可以有向的或无向的，具有不同的语义含义和属性。

3.3

三元组 triple

知识图谱中表示事实的基本单位，由三个元素组成，通常采用主谓宾的结构来描述。它由主语、谓语、宾语三个部分组成，用来表示实体之间的关系或描述实体的属性。通常的三元组采用(h, r, t)进行表示，h为主语或者头实体，t为宾语或者尾实体，r为谓语或者关系，其代表了实体h与实体t之间存在r关系。

3.4

知识图谱 knowledge graph

一种基于图论和语义网络的知识表示模型，它是一种结构化的、语义丰富的知识库，用于描述现实世界中实体之间的关系、属性和概念。

注：这种图形化的数据结构包含了丰富的语义信息，以三元组的形式表示为实体之间的关系，是一种用于组织和表示知识的形式化框架。

3.5

知识推理 knowledge reasoning

利用已知的事实、规则或信息，通过逻辑推理、推断或推理规则等方法，从已有的知识中推导出新的信息或结论的过程，通常包括了现有知识的正确性判断以及未知知识的推理。其通常采用查询三元组的方式进行，即（头节点，关系，？）或（头节点，？，尾节点），在三元组已知两个元素的情况下预测第三个元素。对于需要从结果推导原因的任务（例如故障溯源等），通过构建反关系三元组，将（头

实体，关系，尾实体）转换为（尾实体，反关系，头实体）的相反的三元组，再进行对于现有的反关系三元组（尾实体，反关系，？）的推理，而非直接采用尾节点+关系推理头节点。

4 知识获取、整合和表示

4.1 整体架构

标准制造业产品的全生命周期数据知识推理技术从数据接入开始，对于数据进行预处理以及格式单位上的统一，随后依据数据完成知识图谱的构建，并借助多种算法深度发掘知识图谱中的内容，完成知识推理。

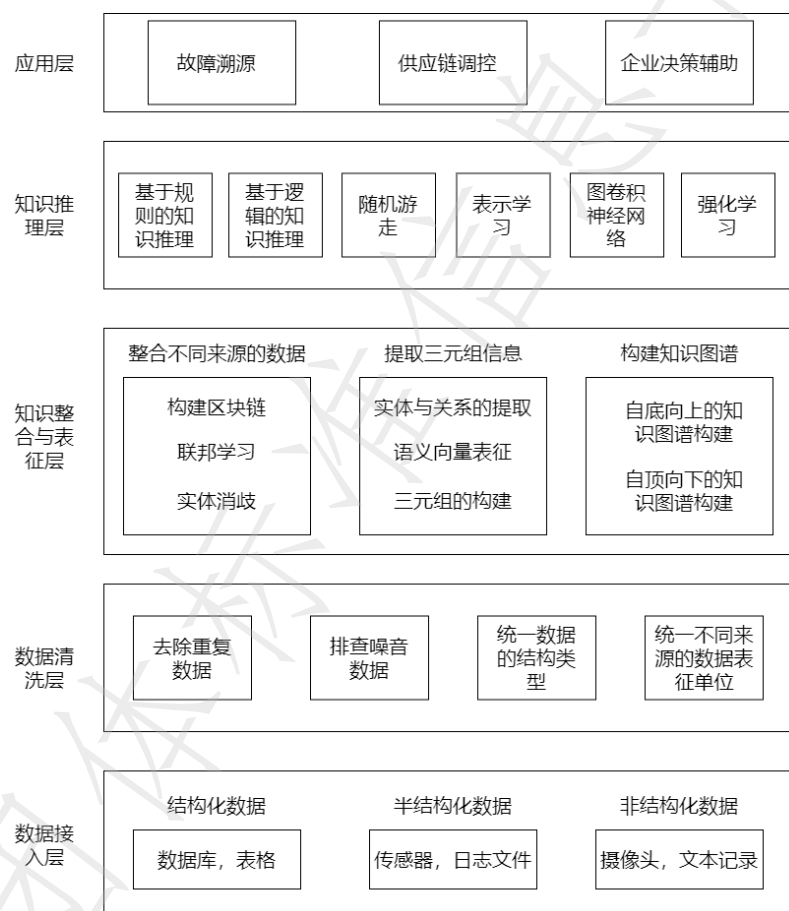


图1 知识推理技术框架图

4.2 数据接入层

数据接入层从产品全生命周期中获取构建知识图谱所需的原始数据，主要包括但不限于生产订单，产品规格，库存记录等直接存储在数据库或者电子表格中的结构化数据，生产过程中传感器、日志文件等半结构化数据，以及监控录像音频文件等非结构化数据。

4.3 数据清洗层

数据清洗层对接入的数据进行预处理和标准化，提升数据质量，确保系统能够有效利用全生命周期的数据。其中功能包括但不限于数据去噪、去重，缺失值填补，数据格式统一等。

4.4 知识整合与表征层

拥有对于不同来源数据的整合能力，并且能够从数据中提取三元组并构建知识图谱。其中包括但不限于对于不同来源的数据的整合，从整合后的数据中提取实体以及关系并构建三元组，以及将实体关系三元组用于构建知识图谱。

4.5 知识推理层

通过基于逻辑或规则的推理，随机游走算法，深度学习等技术，对于现有的知识进行分析，并且推理数据中并未包含的其他知识。

4.6 应用层

基于用户的需求构建所需要的应用，其中包括但不限于故障溯、供应链调控、决策辅助等。

5 算法构建要求

5.1 数据采集

应包含尽可能多样化的数据，更加全面的凸显制造业产品全生命周期之间的联系。数据依据结构化的程度应分为三类。

- 结构化数据：含有固定格式能够被计算机直接处理的数据。通常生产订单，产品规格，库存记录等直接存储在数据库或者电子表格中的数据属于结构化数据。这部分数据可以直接接入到系统中。
- 半结构化数据：具有某种程度上的结构，但不符合传统的表格形式。传感器数据，日志文件等属于半结构化数据，这部分数据需要手动建立数据的提取规则。
- 非结构化数据：非结构化数据是指没有固定格式和明确结构的数据，以自然语言、图像、音频或视频的形式存在。视频监控录像、文本文件、音频文件等属于非结构化数据。这部分数据需要借助文本分析，语音识别等技术进行处理。

5.2 数据清洗和预处理

新采集的数据并不能直接投入应用，应经过数据的清洗和预处理环节，提高知识图谱构建准确度。主要的步骤包括以下步骤：

- a) 数据去重需要保证采集到的数据集中的数据记录不含重复项，对于重复记录的数据进行删除；
- b) 数据去噪是完成对于数据中缺失值的填充，对于异常值的修正，排除数据中的噪音对于模型的影响；
- c) 统一结构类型是将不同来源不同数据结构的数据进行统一，保证后续代码对于数据的处理速度；
- d) 统一数据表征单位是将不同来源中对于同一个数据的表征方式以及单位进行统一。

5.3 整合不同来源的数据

源自不同部门以及系统的数据应使用整合算法进行处理，主要包括以下三种：

- 实体消歧：借助语义向量的相似性以及聚类结果判断两个实体的语义是否一致，融合不同表征的相同实体；
- 构建区块链存储数据：采用去中心化不可篡改的区块链技术存储全部的数据，确保数据的安全性以及可追溯性；
- 联邦学习：在无法提供数据的前提下，多个企业共同训练一个神经网络，共享模型的更新情况，无需将数据统一上传。

5.4 知识表示以及知识图谱的构建

5.4.1 知识图谱的构建应分为三步。

- a) 实体以及关系的提取：识别文本或数据中所包含的实体，并将实体与实体之间存在的关联关系记录。常用的方法包括：
 - 1) 依据人工制定的规则进行提取：依据提前设定好的人工规则与数据进行匹配；
 - 2) 隐马尔科夫链：依据统计的信息判断实体以及关系之间的状态转移概率，从而识别实体和关系的具体类型；
 - 3) 语义向量提取：基于语义信息提取数据文本中的实体，并依据其在向量空间中的表征识别数据之间的关系。

- b) 构建三元组表征：将两个实体以及他们之间的关系构建三元组。
- c) 构建知识图谱：知识图谱由实体，关系，它们的属性以及三元组所构成。依据任务的不同，知识图谱的构建可以分为两类：
- 1) 自底向上的知识图谱构建：首先会从数据中抽取实体关系以及三元组，并自动的填充到知识图谱中，适用于需要自动化的处理大量数据的场景；
 - 2) 自顶向下的知识图谱构建：首先会借助专家知识搭建知识图谱框架，随后从数据中提取符合要求的实体关系以及三元组加入知识图谱的构建，适用于任务需求明确且涉及到的数据类别不多的场景。

5.5 知识推理

应包括判定两个实体之间所存在的关系正误，推断两个无关实体之间所隐含的关联关系，判断在特定关系下实体与其他实体的联系情况等，知识推理能力的强弱代表了制造业知识图谱的智能程度。常见知识推理算法包括以下内容。

- 基于规则的知识推理：人为的设定规则，依据匹配到的规则自动生成新的知识。规则需要手动维护，适用于精度要求较高的简单场景中。
- 基于逻辑的知识推理：考虑常用的谓词逻辑词，将现有的数据转换为逻辑表达式，并自动计算多个逻辑表达式合并后的新知识。但是计算复杂度较高，适用于非结构化的文本数据较多的场景。
- 随机游走：模拟在知识图谱中随机游走，依据生成的路径探索其中节点的隐含关系。算法存在一定的随机性，适用于数据量较小的知识图谱数据集中使用。
- 表示学习：将知识表示为向量的形式，通过向量之间的距离以及相似度实现新知识的推理。适用于大规模的文本，图像等非结构化数据。
- 图卷积神经网络：将知识图谱的节点和边转化为向量或矩阵的形式，对节点及其相邻节点的特征进行聚合和更新，以获取节点的表示。方法能够快速的处理复杂的图结构数据，并且获得较高的推理准确率。
- 强化学习：知识推理将被抽象成为在图上的寻路过程，强化学习训练智能体来实现自动的寻找知识图谱中寻路过程的最佳动作选择策略。该方法主要适用场景为数据变化较大的场景，最终获得的推理模型能够有较强的适应性。

5.6 评价指标

5.6.1 平均排名 (Mean Rank)

$$MR = \frac{1}{|S|} \sum_{i=1}^{|S|} rank_i \cdots \cdots \cdots (1)$$

式中：

S ——三元组集合；

$|S|$ ——三元组集合个数。

$rank_i$ 是指第 i 个三元组的链接预测排名。排名越靠前 $rank$ 越小，求和也就更小，这个指标越小越好，适用于规模较小的数据集。

5.6.2 平均倒数排名 (Mean Reciprocal Ranking)

$$MR = \frac{1}{|S|} \sum_{i=1}^{|S|} \frac{1}{rank_i} \cdots \cdots \cdots (2)$$

式中：

S ——三元组集合；

$|S|$ ——三元组集合个数。

$rank_i$ 是指第 i 个三元组的链接预测排名。排名越靠前 $rank$ 越小，求和也就更大，这个指标越大越好，适用于规模较大的数据集。

5.6.3 前K位命中率 (Hits@K)

$$Hits@K = \frac{1}{|S|} \sum_{i=1}^{|S|} Hits_i \dots\dots\dots (3)$$

式中:

$Hits@K$ ——在前K个返回的结果中的命中率。

$Hits_i$ 代表第*i*次推理前K个将结果是否命中, 若命中 $Hits_i = 1$, 反之 $Hits_i = 0$, K 的取值为1, 3, 10, 该指标越大越好, 适用于一对一关系的推理过程。

5.6.4 前K位召回率 (Recall@K)

$$Recall@K = \frac{1}{|S|} \sum_{i=1}^{|S|} \frac{Positive_i}{Positive_{num}} \dots\dots\dots (4)$$

式中:

$Recall@K$ ——在前K个返回的结果中的召回率;

$Positive_i$ ——第*i*次推理返回的前K个结果中正样本的数量;

$Positive_{num}$ ——正样本的总数。

通常 K 的取值为1, 3, 10, 该指标越大越好, 代表所有准确的条目有多少被检索出来。

5.6.5 前K位准确率 (Precision@K)

$$Precision@K = \frac{1}{|S|} \sum_{i=1}^{|S|} \frac{Positive_i}{K} \dots\dots\dots (5)$$

式中:

$Precision@K$ ——在前K个返回的结果中的准确率;

$Positive_i$ ——第*i*次推理返回的前K个结果中正样本的数量。

通常 K 的取值为1, 3, 10, 该指标越大越好, 代表检索出来的条目有多少是准确的。

6 应用场景

6.1 故障溯源

6.1.1 故障溯源场景需要将生产过程中的相关数据进行汇总, 并且能够实现对于已知故障之间的隐藏关系判断, 其流程如图2所示。

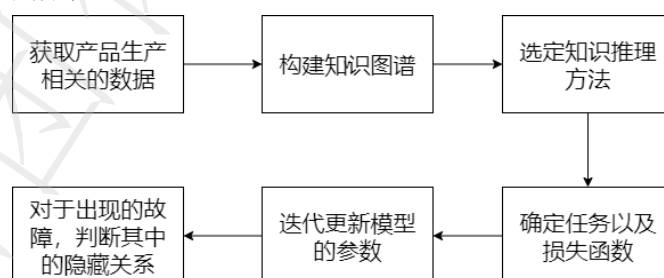


图2 故障溯源流程图

6.1.2 应收集来自传感器、设备监控系统和维护记录等多个数据源的数据, 包括但不限于设备运行状态、温度、振动、压力等参数, 以及维修日志和保养信息。

6.1.3 采用已知的设备知识和历史数据, 构建故障模式的知识库或规则库, 描述设备运行中的各种故障情况和可能的根本原因。

6.1.4 针对现有的需求选定知识推理方法以及损失函数, 构建知识推理模型。模型注重于探索故障节点以及其他节点的关系, 并能够依据输入的故障情况返回关联度较高的其他故障。因此应采用自顶向下的知识图谱构建方法, 精准锁定故障相关的实体, 并且结合图卷积神经网络模型, 用平均排名以及前K项召回率作为评价的指标。

6.1.5 实时收集到的设备数据与知识库进行对比和分析, 以识别潜在的故障模式或异常情况。

6.1.6 结合实时监测数据和已知故障模式，利用知识推理方法预测可能出现的故障，提前发出预警并建议维护措施。

6.2 供应链调控

6.2.1 供应链调控场景需要整理历史订单，市场供给需求变化，为后续的供应关系提供指导以及对于潜在的风险进行预警，其流程如图3所示。

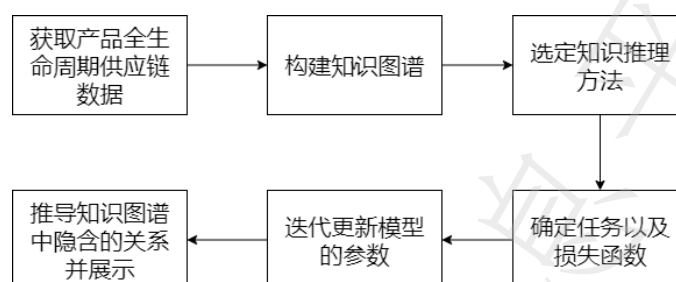


图3 供应链调控流程图

6.2.2 通过知识推理技术解决优化供应链管理、预测需求、降低风险，并提高整体运营效率。

6.2.3 对于历史数据，结合历史订单、销售数据和市场趋势等信息，应建立供应链需求预测模型，可使用知识推理技术，综合考虑市场因素、季节性变化和产品特性，改进需求预测模型，建立供应商评估体系，整合供应商绩效、质量、交付能力等数据。

6.2.4 针对现有的需求选定知识推理方法以及损失函数，构建知识推理模型。模型注重于探索故障节点以及其他节点的关系，能够展示全部种类的节点的关系并且推理原本知识中不存在的信息。因此其更为适合的建模方式是采用自底向上的知识图谱构建方法，借助表示学习来建模多种不同的实体以及关系，并对于节点之间的未知关系，采用前 K 项命中率来进行评估。

6.2.5 应采用知识推理技术分析供应商数据，识别潜在风险和稳定性，辅助供应商选择和管理。

6.2.6 应对预测需求和知识推理结果，优化库存管理策略。

6.3 决策辅助

6.3.1 决策场景需针对不同的决策需求，从知识图谱中提取对应的数据，完成知识推理，为决策者展示实体之间的隐含关系。其流程如图4所示。

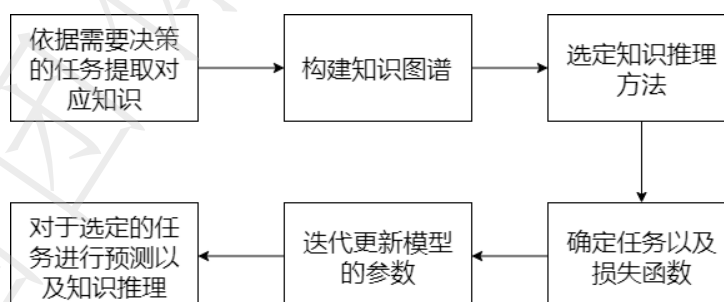


图4 决策辅助流程图

6.3.2 应对不同的任务场景，首先从原本的全知识图谱中提取出来对应的知识来构建一个小型的仅包含所需要内容的知识图谱。

6.3.3 针对现有的需求选定知识推理方法以及损失函数，构建知识推理模型。模型注重于稳定应对不同的任务目标，辅助管理者进行决策。因此其更为适合的建模方式是采用自底向上的知识图谱构建方法，辅以专家知识进行子图的筛选。模型的训练过程通常适用图神经网络进行，并用准确率来衡量模型的具体效果。此外还需要考虑模型的鲁棒性以及面对不同任务的可扩展性。

6.3.4 应采用知识推理技术帮助制造业决策者优化生产流程、提高效率、降低成本，有效应对市场变化。

6.3.5 应采用知识推理技术分析生产数据、工艺参数和设备状态，识别生产流程中的潜在问题和瓶颈，

提出优化建议。

6.3.6 结合知识图谱的知识可视化能力提供决策支持，辅助制造业决策者快速做出合理决策，帮助企业更好地理解生产过程和各环节之间的关系，提高预测准确性，降低生产成本，优化供应链管理，从而提高企业运营效率和竞争力。
