

装备保障性验证知识图谱构建方法研究

刘晨光¹, 李星新¹, 于永利¹, 孙也尊²

(1.陆军工程大学石家庄校区, 河北 石家庄 050051;

2.中国人民解放军32382部队, 北京 100000)

✉17370228702@163.com; lxx_1226@sina.com; yu_yongli@263.net.cn; 461149680@qq.com



摘要: 装备保障性验证是控制装备保障性达到目标要求的有效手段, 目前, 装备保障性验证领域积累了很多技术方法以及海量异构的数据。知识图谱具有对知识、数据进行梳理从而进一步实现机器智能的作用。构建装备保障性验证领域知识图谱也将是装备保障性领域迈向智能化的开端。首先简要介绍装备保障性验证发展现状, 其次概括知识图谱构建的技术方法和研究热点, 然后提出装备保障性验证领域知识图谱构建方法以及逻辑和技术流程, 最后对未来发展前景做了展望。

关键词: 装备保障性; 异构数据; 知识图谱; 本体; 知识抽取

中图分类号: TP399 **文献标识码:** A

Research on Construction Method of Knowledge Graph for Equipment Supportability Verification

LIU Chenguang¹, LI Xingxin¹, YU Yongli¹, SUN Yezun²

(1.Shijiazhuang Campus of Army Engineering University, Shijiazhuang 050051, China;

2.Unit 32382 of PLA, Beijing 100000, China)

✉17370228702@163.com; lxx_1226@sina.com; yu_yongli@263.net.cn; 461149680@qq.com

Abstract: Equipment supportability verification is an effective means to control equipment supportability to meet target requirements. At present, the field of equipment support verification has accumulated many technical methods and massive heterogeneous data. Knowledge graph has functions for sorting knowledge and data to further realize machine intelligence. Building knowledge graph for equipment supportability verification marks the initial effort of intelligence in the field of equipment supportability. This paper firstly introduces the development status of equipment supportability verification. Then, it summarizes the technical methods and research focuses of knowledge graph construction. Thirdly, it proposes the method as a logic and the technical process of knowledge graph construction in the field of equipment supportability. Finally, it depicts the future development.

Keywords: equipment supportability; heterogeneous data; knowledge graph; ontology modeling; knowledge extraction; knowledge fusion

1 引言(Introduction)

保障性是装备的设计特性和计划的保障资源满足平时战备和战时利用率要求的能力^[1], 保障性的概念分为三个层次, 一是装备的设计特性, 这是在装备设计阶段赋予装备天生的特性; 二是计划的保障资源, 这是使装备得到有效保障的物质基础; 三是二者协调配合满足装备战备完好性的能力, 这是需要装备做到“好保障”与“保障好”相配合, 使装备达到平时战备和战时使用要求^[2]。

近年来, 随着我军大批新型武器装备列装部队以来, 装备的复杂程度和信息化程度越来越高, 对装备保障性水平提出了更新更高的要求。在借鉴世界军事强国发展经验的基础上, 我军对装备全寿命周期内装备保障性的研究越来越重视, 取得的理论成果、技术方法越来越全面, 在此基础上,

进一步构建装备保障性验证领域知识图谱, 将对理论方法指导部队实际工作产生重要积极的推动。

2012年谷歌首次提出知识图谱的概念, 知识图谱在语义搜索、问答系统、智能客服、个性化推荐、辅助决策等应用中占有重要作用, 金融、医疗、司法、公安等许多行业都在探索构建垂直领域知识图谱, 致力于实现行业智能化水平^[3]。当前较为典型的开放领域知识图谱有: DBpedia^[4]、YAGO^[5]、Probase^[6]、BableNet^[7]、Zhishi.me^[8]等。但是对于军队系统, 我们将理论成果赋能试验基地、基层部队的效果还比较差, 开展工作的方式比较传统、严重制约专业领域知识向下传递、实践经验与数据向上反馈的过程。构建保障性验证领域知识图谱将是专业领域知识赋能部队, 提升部队智能化水平的重要基础。

2 相关工作(Related work)

2.1 保障性验证的主要研究内容

装备保障性验证的研究内容与保障性定义的内涵是相匹配的。

(1)对保障性设计特性验证方法的研究。针对装备系统故障率、平均故障间隔、保持或恢复到规定状态的能力等问题,依据可靠性维修性保障性(RMS)指标要求,构建试验模型、确定试验项目、优化试验流程,科学检验装备系统的保障性设计与工艺缺陷^[9]。

(2)保障资源规划的验证研究。针对配套保障设备的完备情况、符合通用化系列化组合化要求的程度、备附件的供应满足率、技术资料的完整性和适用性、保障工作量以及综合保障各要素之间匹配性协调性等问题,建立保障资源配套试验评价标准,构建试验模型,确定试验项目,优化试验流程,全面考核装备保障资源的利用率和充足程度,以及保障系统的能力是否和装备战备完好性要求相适应^[10]。

(3)战备完好性验证方法研究。主要考虑装备系统是否正常工作和使用,以及能否在预期作战环境中按照预定的强度执行作战任务等要求,建立装备系统可用度试验评价标准,构建试验模型,优化试验流程,科学考核装备系统处于能够使用并能够执行任务的状态的程度^[11]。

2.2 知识图谱定义

知识图谱是AI+知识的一种表现形式,以类脑神经的图结构将大数据背景下的海量知识相互关联实现进一步智能化的技术方法。知识图谱可视化图结构中的节点代表现实世界的实体(如,人、地点、机构等)或者抽象概念(如,装备保障性、装备维修性等);两个节点之间的边代表实体的属性(如,人名、地名)或者实体之间的关系(如,保障、维修等)^[12]。知识图谱是在语义网络的基础上进一步发展而来的产物,是将知识联系存储成一个大型知识库。知识图谱所具有的推理能力可以帮助人类发现新的知识和关系,与关系型数据库相比,知识图谱的类人脑的知识组织结构有助于智能化的处理和理解知识,进一步提高机器智能^[13]。

2.3 知识图谱架构

2.3.1 逻辑层面架构

在逻辑层面,知识图谱通常可以划分为数据层和模式层两个层次。数据层主要是将领域中的多源异构数据转换为三元组的形式,通过一系列三元组表达领域知识或者常识知识,比如用(实体1,关系,实体2)、(实体、属性,属性值)这样的三元组来表达事实,从而实现对知识进行存储。模式层是在数据层的基础上构建的,领域知识图谱构建过程通常通过构建领域本体(Ontology)来规范数据层。本体规范了领域核心结构,具有较强的结构层次,有效减少冗余^[14],如图1所示。

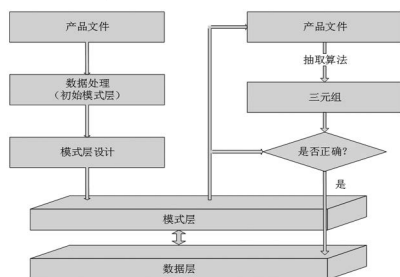


图1 逻辑层架构示意图

Fig.1 Schematic diagram of logical layer architecture

知识图谱构建方式分为自底向上(bottom-up)和自顶向下

(top-down)两种。“自底向上”的方式是指从数据层入手,通过对现实世界海量庞杂的知识数据进行初步分析,按照特定领域数据类型分布,以及数据特点进而构建与之相匹配的知识图谱模式层,即先构建底层数据层进而推演出顶层模式层^[15]。由于通用领域知识图谱主要强调知识的广度以及知识之间浅层的关联关系,因此一般采用自底向上的构建方式,如Google的Knowledge Vault。另外,在领域知识图谱过程中,部分知识体系不完整的数据也需要采用自底向上的构建方式;“自顶向下”的方式是在领域知识体系成熟完备的条件下首先进行知识图谱模式层的构建,进而补全相应数据层,这种构建方式需要领域专家早期介入,利用领域内固有的知识体系或者新构建权威的知识体系,并在此基础上梳理可定义模式的数据。由于垂直领域知识图谱对领域知识准确性以及领域知识之间深层次关系要求较高,因此一般采用自顶向下构建方式,如医学领域知识图谱。本文构建装备保障性验证领域知识图谱主要采用自顶向下的方式构建。

2.3.2 技术层面架构

垂直领域知识图谱是相对Wikidata、Yago等通用知识图谱而言的,垂直领域知识图谱的知识复杂度更高、知识来源更加复杂、知识数据的多元异构性更强。现实世界的知识资源海量庞杂,由于通用知识图谱要求知识覆盖面庞大,因此通用知识图谱构建通常以知识数据为牵引采用自底向上的构建方式。而垂直领域知识图谱一般面向不同的特定领域和不同的数据模式,因此现在没有统一的构建标准和规范。但是,一般构建过程可以分为六个阶段,即知识建模、知识存储、知识抽取、知识融合、知识计算、知识应用,如图2所示。

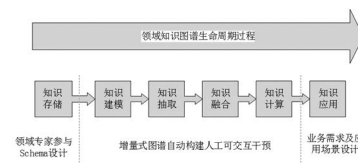


图2 知识图谱生命周期

Fig.2 Life cycle of knowledge Graph

根据图中的各个构建部分,其技术层面架构如图3所示。

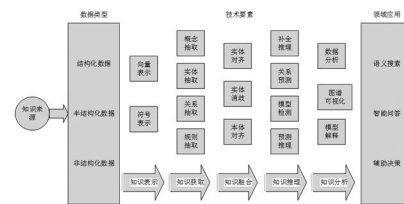


图3 技术层架构

Fig.3 Technology layer architecture

3 面向装备保障性验证的知识图谱构建 (Construction of knowledge graph for equipment supportability verification)

3.1 装备保障性验证知识图谱内涵

装备保障性验证知识图谱是知识图谱在装备保障领域的延伸,是图结构的装备保障性验证语义知识库,通过形式化描述装备保障性验证领域概念、实体、属性,以及之间关系的方式,使领域内概念、实体、属性相互连接,构成装备保障性验证领域知识的立体网状结构图,进而在装备保障领域实现知识图谱的语义搜索、智能问答、辅助决策等相关功能。

装备保障性验证领域知识可以采用资源描述框架(Resource Description Framework, RDF)表示为三元组形

式, RDF本质是一种数据模型(Data Model), RDF假设所有语义都可以以三元组形式进行表示, 复杂语义可以由n个三元组组合表示, 它提供了一种统一的描述实体、概念、属性的标准, 形式上表示为SPO三元组形式, 即“<s(主语), p(谓语), o(宾语)>”或者“<(对象), (属性), (值)>”, 例如可以用如下三元组表示装备技术准备时间的基本信息:

<装备技术准备时间, 类型, 保障性参数>
 <装备技术准备时间, 验证方法, 半实物仿真>
 <装备技术准备时间, 父类, 时间>
 <半实物仿真, 父类, 试验>

3.2 装备保障性验证知识图谱特点

知识图谱是人工智能范畴的新兴领域, 装备保障性验证知识图谱是开创性的将知识图谱技术引入到装备保障性验证领域, 它不同于知识库或者专家系统的功能, 不仅仅是对装备保障性验证领域的知识总结, 而是实现装备保障性领域更具智能化的重要探索, 需要针对保障性验证的特点, 对概念、实体、关系进行梳理扩展, 具有以下特点:

(1)装备保障性验证知识图谱是一项装备保障领域知识工程, 是在对装备保障性验证领域知识形式化表示的基础上, 进行知识抽取、知识融合、知识众包、知识表示、知识推理以及可视化等工作, 并最终基于保障性验证知识图谱实现语义搜索、智能问答、辅助决策等面向部队使用的全寿命过程。

(2)装备保障性验证知识图谱描述的关系包括人、装备、时间、空间、工具、技术等之间复杂交错的关系, 复杂关系交叉使得事件逻辑性和复杂性大大增加, 在这种专业领域背景下, 将发挥知识图谱技术的特点, 提升装备保障性验证工作的智能化水平。

(3)装备保障性验证知识图谱需要描述事实型知识和过程型知识, 事实型知识指的是有装备保障领域专家公认的装备保障术语、名词、框架等反映装备保障性验证目的与状态的知识; 过程型知识指的是描述装备保障性验证系统过程的知识, 具有较强的针对性, 体现装备保障性验证的专业特点, 利用知识图谱技术可以较好的建立知识之间的联系。

3.3 装备保障性验证知识图谱构建技术体系

根据装备保障性验证领域知识特点, 综合领域知识图谱构建一般方式, 采用“自顶向下”的构建方法, 首先构建领域的模式层, 采用本体建模的方式构建装备保障性验证本体模型。数据层为装备保障性领域实体和相关属性, 实现保障性验证知识抽取、融合、推理、补全。

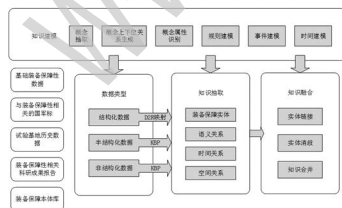


图4 装备保障性验证知识图谱构建技术流程

Fig.4 Technical process of constructing knowledge graph of equipment supportability verification

(1)装备保障性验证本体建模

本体的概念最初起源于哲学, 从20世纪90年代开始, 计算机领域特别是人工智能、知识工程领域开始广泛引入本体概念, 形成各个应用领域内通用标准化的领域概念, 成为知识表示、加工、应用的基础。现在本体的形式化定义主要有OWA形式化定义、Guarion形式化定义和KAON形式化定

义, 其中OWA形式化定义应用最为广泛, 本文采用这种方法进行装备保障性验证领域本体建模。OWA形式化定义是一种基于Web注释的本体描述方法, 其形式化定义为:

$$O = \langle C, A^C, R, A^R, H, X \rangle$$

其中, C 是领域概念集; A^C 是概念的属性集; R 是关系集; A^R 是每个关系属性集; H 表示概念层次; X 表示公理集。

装备保障性验证本体建模实际上就是规范化描述装备保障性验证系统过程, 将装备保障性验证作为一个本体进行概念抽象和分析, 并将每一层概念作为一个本体, 最终依据抽象分析的结果确定模型框架。

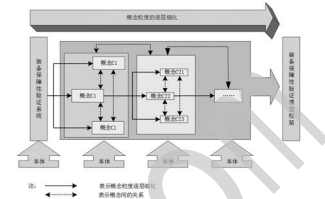


图5 本体建模分析思路

Fig.5 Ontology modeling and analysis

基于本体的保障性验证系统分析的核心在于: 在OWA形式化规范下逐层确定装备保障系统概念和概念之间的关系。首先, 将整个装备保障性验证系统作为一个本体, 并基于OWA进行分析, 界定装备保障性验证系统的顶层概念及概念之间的关系。然后, 将每一个概念分别作为一个本体, 如此逐层分析, 直至确定研究目的下所定义的最小细粒度概念。在概念分析的过程中会逐层体现所需建立的模型。

(2)多源异构的装备保障性验证知识抽取

多源异构装备保障性实体抽取和实体关系构建包括装备保障性验证知识抽取、关系链接、关系推理等。保障性验证领域知识的来源包括装备保障本体库、基础装备保障性数据、与装备保障性相关的国军标、装备保障性相关的科研项目成果(报告、模型)、试验基地的历史数据等。可从中提取装备保障性验证相关概念、装备保障性验证实体、装备保障性验证相关属性、语义关系、空间关系、时间关系。

装备保障性验证相关概念主要源于装备保障性验证本体的构建, 如保障性使用参数、战备完好性参数、任务成功性参数等都是概念范畴, 实体是概念的实例, 含有相应的具体的属性, 如能执行任务率(MC)、再次出动时间(TAT)、使用可用度(Ao)等都属于装备保障性验证实体。装备保障性验证关系主要有语义关系、时间关系、空间关系。语义关系包括“IsA”关系、相似关系、互斥关系。“IsA”关系是一种行为关系, 既包括概念之间的父子关系(如“保障资源”与“保障设备”), 也包括概念与实体的(如“保障设备”与“扳手”)的实例关系; 相似关系包括用来描述意思上相似的同级保障性验证领域概念, 如(“发射准备时间”与“技术准备时间”); 互斥关系用来描述同级保障性验证领域关系之间以及相同性质实例之间的互斥关系, 如(“可靠度函数R(t)”与“累计故障分布函数F(t)”)。

时间关系主要描述在装备寿命周期各阶段装备保障性验证所处的不同状态以及任务属性。对时间关系的形式化描述可分为两类, 一类是装备保障性验证事件、过程等实体之间的时变关系描述, 如“人员→项目”“参数→特性”, 非结构化的保障性验证事件, 如事件: “在装备设计阶段, 装备设计人员需要考虑加油口盖泡沫功能”可对抽取得到的三元组“设计人—设计—加油口盖泡沫功能”添加时间标签拓

展为四元组,用来描述时间;第二类是空间特征的时间变化属性。可用于实体和关系的更新,如装备进行长途奔袭过程中,装备的空间属性值以及装备状态属性值是随时间而变化的,这也是未来构建动态知识图谱的时间轴基础。

(3)装备保障性验证知识融合

经过知识抽取得到的知识可以划分为本体层(抽象知识)和实例层(具体事实)。一方面,虽然通过本体建模的方式构建了装备保障性验证领域的本体模型,但是事实上,无法构建出一个能够包含全方面的统一本体,因为面向不同应用场景以及面向不同的用户使得不同本体之间有不同侧重点和差异,此外,过于庞大的本体也难以维护和使用,在工程上有很大困难;另一方面,大量实例之间也存在异构性问题,例如同名实例可能指代不同实体,不同名实例可能指代同一实体,而且知识之间可能存在冗余和错误的情况。因此,知识融合需要解决本体层和实例层异构问题,以提高知识质量和精度。

针对本体层,本文采用基于术语的本体映射方法解决其异构的问题,从本体中术语的相似性来寻找本体映射。如图描述了本体映射的基本过程。

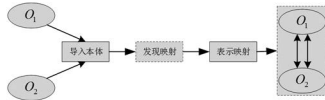


图6 本体映射生成过程

Fig.6 Ontology mapping generation process

在规范字符串的基础上,通过度量字符串的相似程度判断本体的映射关系,可以使用计算字符串汉明距离、子串相似度、编辑距离和路径距离的方法度量字符串相似度。计算方法如下:对于导入的任意两个本体字符串 s 和 t ,他们的汉明距离为:

$$\delta(s, t) = 1 - \frac{(\sum_{i=1}^{\min(|s|, |t|)} s[i] \neq t[i]) + ||s| - |t||}{\max(|s|, |t|)}$$

若存在两个字符串 p 和 q 使得 $s = p + t + q$ 或 $t = p + s + q$,则称 t 是 s 的子串或 s 是 t 的子串,令 x 为 s 和 t 的最大公共子串,则 s 和 t 的子串相似度为:

$$\delta(s, t) = \frac{2|x|}{|s| + |t|}$$

若给定一个字符串操作集合 op 和一个代价函数 w ,对于任意一对字符串 s 和 t ,存在将 s 转化为 t 的操作序列集合,则两字符串的编辑距离为:

$$\delta(s, t) = \min \sum_{i=1}^n w_{op_i}, \text{ 且 } op_n(\dots op_1(s)) = t$$

若给定两个字符串序列 $\langle s_i \rangle_{i=1}^n$ 和 $\langle s'_j \rangle_{j=1}^m$,则路径距离计算如下:

$$\delta(\langle s_i \rangle_{i=1}^n, \langle s'_j \rangle_{j=1}^m) = \lambda \cdot \delta'(s_1, s'_1) + (1 - \lambda) \cdot \delta(\langle s_i \rangle_{i=2}^n, \langle s'_j \rangle_{j=2}^m)$$

其中, δ' 是某字符串度量函数, $\lambda \in [0, 1]$ 。

针对实例层,实例层的融合匹配与本体层具有相似之处,但是由于知识图谱的实例层往往数据规模庞大,因此,在匹配过程中需要考虑时间复杂度和空间复杂度问题。结合装备保障性验证领域特点,采用基于规则的实例匹配方法。

4 装备保障性验证知识图谱应用(Application of knowledge map of equipment supportability verification)

知识图谱的应用需要将知识图谱特有的应用形态与特定的应用场景以及应用对象相结合。通用领域知识图谱在语义搜

索、智能问答、可视化决策等功能的应用十分典型,针对装备保障性验证知识图谱的应用需要基于数据特点以及业务场景进行优化,可以分为面向人应用以及面向智能平台的应用。

4.1 基于装备保障性验证知识图谱的智能知识服务

装备保障性验证知识图谱以可视化的图结构存储专业领域的相关知识,增强了领域知识的结构和语义,使得知识成为人可视而且机器可处理的形式。基于装备保障性验证知识图谱的语义搜索与传统的装备保障性知识库搜索相比可以处理粒度更细、结构更强的语义数据。实现以装备保障实体的形式在统一视图里进行管控。

4.2 基于装备保障性知识图谱的智能问答

智能问答系统首先要经过语义理解和知识匹配两个过程,智能搜索系统同样要经过这两个过程,相比较而言智能问答系统可以看作是更高级的信息检索方式。基于知识图谱的智能问答相比较传统基于字符匹配的问答具有更准确的答案反馈,更智能的问答交互。

面向从事装备保障性验证相关的工作人员,以及进行装备保障相关培训的学员,可以以基于自然语言的问句输入,将得到快速、准确的信息反馈。问答系统可以实现问句的智能理解与分解,逐步迭代得到最佳匹配的答案,再将答案进行整合反馈。面向装备保障性验证领域知识的智能问答将是保障装备性知识图谱的中重要应用方向,保证在相关工作,以及学习中智能化的专业知识辅助。

4.3 基于装备保障性知识图谱的决策支持分析

在装备保障性验证工作中,往往涉及多维数据的综合分析与评价,其中数据挖掘分析基本数据下的隐含表现并结合装备保障性验证工作的专业领域知识的过程是一项复杂困难的工作。基于装备保障性验证知识图谱可以通过相关的数据挖掘算法、图计算方法,采用智能辅助的方式实现在专业分析的过程中进行实体的关联关系分析、关联数据分析等功能。实现在时间和空间维度离散分布的情况下对数据、知识、相关资料进行整合处理,辅助相关人员的决策分析,提高决策的准确性和效率,提高装备保障领域工作的智能化信息化水平。

5 结论(Conclusion)

本文剖析了装备保障性验证知识图谱的概念和内涵,重点研究了基于多源异构数据的装备保障性验证知识图谱构建的技术体系和方法。目前,在更强大的计算能力、更先进的算法、爆炸式数据等诸多因素的共同作用下,人工智能领域蓬勃发展,取得许多成熟技术成果,将人工智能赋能军队专业技术性领域提高部队开展工作的智能化水平成为一种必然趋势,构建装备保障性验证知识图谱是探索装备保障领域实现智能化的重要基础。

知识图谱提供了一种从数据中抽取结构化知识并利用图分析进行关系挖掘的手段,具有广阔的应用前景,也充满技术挑战。未来需要进一步提高装备保障性验证知识图谱自动化构建能力、多源异构数据抽取准确性和关系抽取丰富度,并且保障性验证知识质量评价以及动态更新等重要问题还进一步探索研究。

参考文献(References)

- [1] GJB 451A-2005.可靠性维修保障性术语[S].中国人民解放军总装备部批准,2005:2.
- [2] 詹姆斯·琼斯.总装备部技术基础管理中心,译.QRMS-46.

(下转第4页)