

doi: 10.7690/bgzdh.2023.02.002

## 单兵装备人机工效知识图谱构建技术

陈 思<sup>1</sup>, 王亚平<sup>1</sup>, 王新蕊<sup>2</sup>, 胡灿灿<sup>1</sup>

(1. 南京理工大学机械工程学院, 南京 210094; 2. 中国兵器工业第 208 研究所 3 室, 北京 102202)

**摘要:** 为改善单兵装备人机工效领域知识共享困难、利用效率低的现状, 提出基于“一致性检验”与“语料评价”的骨架法应用策略。构建单兵装备人机工效领域知识本体, 对本体网络进行质量评估与优化; 搭建基于“半指针-半标注”抽取思路的联合抽取模型, 进行领域知识文本中重叠三元组的自动化信息抽取, F1-score 达到 0.82; 基于 Neo4j 构建单兵装备人机工效知识图谱, 对图谱的应用方式进行讨论。结果表明: 以主客观评价结合的骨架法应用策略能够提高知识本体构建效率, 达到较好的语料数据覆盖效果; 单兵装备人机工效知识图谱能为领域知识的共享与重用提供数据支撑与技术手段。

**关键词:** 单兵装备; 人机工效; 知识图谱; 本体建模; 骨架法; 联合抽取

**中图分类号:** TJ201 **文献标志码:** A

## Construction Techodogy of Knowledge Map in Individual Equipment Ergonomics

Chen Si<sup>1</sup>, Wang Yaping<sup>1</sup>, Wang Xinrui<sup>2</sup>, Hu Cancan<sup>1</sup>

(1. School of Mechanical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China;

2. No. 3 Research Room, No. 208 Research Institute of China Ordnance Industries, Beijing 102202, China)

**Abstract:** In order to improve the current situation of difficult knowledge sharing and low utilization efficiency in the field of ergonomics of individual equipment, the application strategy of skeleton method based on “consistency check” and “corpus evaluation” is proposed. Construct the ergonomic domain knowledge ontology of individual equipment, evaluate and optimize the quality of ontology network, build a joint extraction model based on the extraction idea of “half pointer and half label”, and carry out automatic information extraction of overlapping triples in domain knowledge texts, with F1-score reaching 0.82. Based on Neo4j, the knowledge map of ergonomics for individual soldier equipment is constructed, and the application mode of the map is discussed. The results show that the application strategy of the skeleton method based on the combination of subjective and objective evaluation can improve the efficiency of knowledge ontology construction and achieve better corpus data coverage. The ergonomics knowledge map of individual equipment can provide data support and technical means for the sharing and reuse of domain knowledge.

**Keywords:** individual equipment; ergonomics; knowledge map; ontology modeling; skeleton method; joint extraction

### 0 引言

为应对复杂多变的战场环境, 满足信息化战场的作战需求, 以士兵为平台的“单兵系统”包含了多种武器装备, 单兵系统总体及单体装备的人机工效越来越为人们重视。国内外学者对单兵装备人机工效开展了大量研究, 领域知识涉及装备学、军需技术、工程设计、人机工程学、服装工程学、医学及数值仿真分析等多个学科。由于单兵装备种类繁多, 其人机工效概念类型繁多, 且联接关系复杂; 目前还未形成完整的知识体系, 知识散落于国军标、学术论文、研究报告以及著作中, 导致领域知识共享困难, 资源利用效率较低。为增强单兵装备人机工效领域内知识的共享能力, 提高知识资源的利用效率, 有必要对单兵装备人机工效领域内的知

识共享方式与需求进行分析, 建立单兵装备人机工效设计知识本体, 采用深度学习对领域知识来源进行联合信息抽取, 构建单兵装备人机工效知识图谱。

为在数据层面解决知识共享困难的问题, 2012年谷歌提出了一种用于提高信息检索效率的形式化语义知识描述框架, 即知识图谱<sup>[1]</sup>。知识图谱是事物关系的可计算模型, 能从数据中识别、发现和推断事物与概念之间的复杂关系。其相关技术主要包含知识图谱构建与图谱应用 2 大方面, 其中前者的主要研究内容包括本体构建与知识抽取 2 方面。

目前知识本体构建方法按数据流向主要分为自顶而下、自底而上以及综合法(自顶向下与自底向上相结合)等 3 类。早期本体构建的方法以人工方式的“自顶而下”方式为主<sup>[1]</sup>, 其中七步法由于通用性

收稿日期: 2022-10-06; 修回日期: 2022-11-28

作者简介: 陈 思(1998—), 男, 河南人, 硕士, 从事单兵装备人机工效、知识工程、专家系统研究。E-mail: ChenSi11011@163.com。

强、技术成熟等优点，在领域知识建模中得到了广泛应用<sup>[2]</sup>。随着机器学习技术的发展，出现了基于统计或基于语义 2 种手段的自动化本体构建方法，即通过基于统计或基于语义 2 种手段，实现了基于数据集“自底而上”地自动化本体构建，但目前自动化本体构建仍面临更新困难、关系消歧和本体消歧等技术难题<sup>[3]</sup>，难以在工程实际中应用。

知识抽取是构建大规模知识图谱的重要环节。三元组抽取模型最初采用 pipeline 式抽取方法<sup>[4-8]</sup>，其串联式的子任务分配导致抽取模型存在误差累计传播的缺陷。因此有学者提出“实体-关系联合学习”的抽取概念，包括基于特征(规则)的抽取模型<sup>[9-12]</sup>以及基于神经网络的模型<sup>[13-17]</sup>。但上述模型难以胜任存在三元组重合现象等语义复杂度较高的抽取任务。Zeng 等<sup>[16]</sup>提出了“三元组”数据的 3 种重叠模式，并将其引入含有复制机制的 seq2seq 模型中；Fu 等<sup>[17]</sup>提出了一种基于图卷积网络的解决方法。但这 2 种方法仍将实体关系作为离散标签进行处理，忽略了三元组内部的依赖关系，导致模型学习困难。因此，将三元组作为整体进行建模<sup>[15]</sup>成为解决复杂语义关系抽取中模型设计的可行思路，如 Wei 等<sup>[18]</sup>提出了一种针对复杂语义的抽取方法。

笔者引入基于语料数据的本体质量评价及迭代

优化框架，采用骨架法应用策略建立本体网络，结合领域语料数据，构建了单兵装备人机工效领域本体，在保证本体任务明确的前提下实现了本体对语料数据较高的数据覆盖率；针对单兵装备人机工效知识片段存在的三元组重合现象，搭建了基于“半指针-半标注”抽取思路的联合抽取模型，实现了领域知识文本中重叠三元组的自动化信息抽取。为单兵装备人机工效领域知识的共享与重用提供了理论与数据基础，有助于单兵装备人机工效设计与评估工作的有效开展。

## 1 单兵装备人机工效领域本体构建

### 1.1 骨架法应用策略

常规的领域知识本体构建，首先要明确知识本体的任务及其表示范围，在具有明确方向的前提下，再进行资料收集、本体类设计及属性设计等环节，构建满足领域知识表示需求的本体网络。然而在本体需求模糊、知识体系架构复杂但知识来源固定的情况下，常规构建流程将难以实施。为适应单兵装备人机工效知识领域现状，笔者基于七步法的方法思路框架<sup>[19]</sup>和软件开发领域中的迭代优化模型<sup>[20]</sup>，提出了一种基于“一致性检验”与“语料评价”的骨架法应用策略，如图 1 所示。

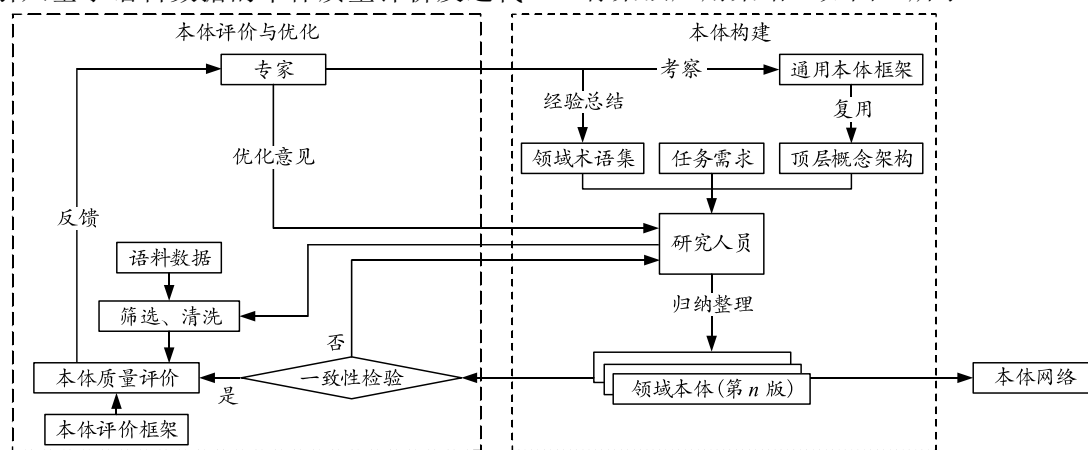


图 1 骨架法应用策略

首先由专家分析及确定“术语-关系”上层框架，基于任务需求和领域特点，进行初步本体网络建设。为确定领域本体的顶层架构，基于领域术语集及本体任务对现有成熟本体的上层架构进行复用调研，不对本体网络进行具体细化，得到本体初步框架。

然后进入本体评价阶段。首先进行一致性检验，若存在缺陷则将本体交由研究人员进行调整；通过一致性检验则进入下一步的评价阶段。评价阶段引入语料数据(部分具有领域知识代表性特征的标注

数据，数据量视领域数据集大小及概念结构复杂度确定)，采用基于语料数据的质量评价框架<sup>[21]</sup>对本体网络的质量进行评价；专家基于质量评价结果，对本体进行结构性及功能性评估并提出改进意见。

研究人员根据专家意见实施对本体网络的优化和评价语料的筛选清洗，再次进行本体评价。循环执行上述操作，直至满足主客观评价要求及功能需求，输出最终本体网络。

笔者提出的骨架法应用策略明确了在构建过程

中的本体演进流程，在循环优化中采用了主客观结合的本体评价与验证流程，可通过在具有代表性的小部分数据集上进行快速迭代优化的方式短时间内得到任务与数据集兼容性较好的本体网络。此应用策略也适用于本体后期的使用及维护阶段中本体的迭代更新，用户可直接将反馈意见交由专家处理，根据专家意见对本体进行评估。

### 1.2 单兵装备人机工效领域知识本体构建

为保证知识图谱的科学性和易用性，笔者采用自顶而下的本体构建方式，从顶层概念向下细分构建知识本体。单兵装备人机工效领域涉及大量的人体解剖结构、生理现象、人体活动等概念，故采用一体化医学语言系统 (unified medical language system, UMLS)<sup>[22]</sup>。

在 UMLS 的顶层概念架构基础上进行领域本体构建，根据单兵装备人机工效设计与评估的使用需求，将领域内的所有概念和术语分为“物理实体”“概念实体”和“事件”3 大类，将关系分为“携行装配相关关系”“功能性能相关关系”“人机工效设计相关关系”和“人机工效试验相关关系”4 大类。

#### 1.2.1 实体建模

根据人机工效的“人-机-环”3 大要素，笔者定义物理实体为人体、装备、试验设备等，概念实体为人机工效设计准则、设计案例、评价方法等，事件为单兵装备作战环境、作战任务等。在本体构建软件 protégé 中的模型结构如表 1 所示。

表 1 本体类架构

一级父类	二级父类	一级子类	二级子类
		人体	-
	人机工效物理实体	人机工效无机实体	装备 实验设备
		人体特性	生物力学特性 认知反应特性
		人机工效设计实例	-
		人机工效设计要求	-
owl:Thing	人机工效概念实体	人机工效评价方法	人机工效实物评价 人机工效半实物 评价 人机工效数值仿真
		装备特性	-
	人机工效事件	作战任务 作战环境	评价指标 实验环境

#### 1.2.2 本体关系建模

遵循“装备”→“人体”→“设计准则”→“评估方法”→“试验设备”的基本思路，为保证本体及其关系对领域知识语义表示的完整性和正确性，

笔者采用“上层框架-关系数据聚类”的方法进行本体网络的概念关系构建。单兵装备人机工效设计及评价数据集中的语义关系，进行人工聚类并优化后得到：携行装配相关关系 2 种、功能性能相关关系 5 种、人机工效设计相关关系 7 种以及人机工效试验相关关系 6 种，本体关系架构如表 2 所示。

表 2 关系架构

一级关系	二级关系	细分关系
		依赖设计要求 相关人体部位 相关装备 相关部位 设计基准 设计装备 设计要求
	人机工效设计相关关系	评价指标 实验环境 执行试验 测量特性 试验任务 试验设备
owl:topObjectProperty	人机工效试验相关关系	人体特性 影响特性 环境因素影响 装备特性 装备特性影响
	功能性能相关关系	依赖装备 关联部位
	携行装配相关关系	

以装备与人体的关系为例，由于单兵装备均为人体直接携行使用，因此可建立如下三元组：{装备，携行部位，人体部位}，{装备，关联特性，人体特性}。

### 1.3 本体质量评价与改进

笔者采用基于语料数据的本体评价法对初步构建的本体进行了评价。基于语料数据的评价方法<sup>[23]</sup>通过考察本体对语料的术语覆盖率(CMM)、关系覆盖率(PCM)、关系扩充率(PEM)和实例扩充率(IEM)4 个参数，来衡量构建本体的质量及其对领域中语义数据的表示能力，计算方法如下：

$$CMM(O) = \sum_{o \in C[O]} \sum_{t \in T} K(o, t) / |T|; \quad (1)$$

$$K(o, t) = \begin{cases} 1: & \text{if (lable}(i) = t) \\ 0.5: & \text{if (lable}(i) \text{ contain}(t)) \end{cases}; \quad (2)$$

$$PCM(O) = |tp| / |cp|; \quad (3)$$

$$PEM[O] = |ap - cp| / |op|; \quad (4)$$

$$IEM[O] = |ai| / |ci|. \quad (5)$$

式中：C[O]和 T 分别表示本体和预料数据中的领域术语集合；K(o, t)表示本体术语与语料中术语的匹配度指数，根据匹配程度选取 1、0.5 或 0；cp 表示

语料中所有的概念关系； $tp$  表示本体覆盖到语料数据中的关系数量； $ap$  表示本体中有定义且能在语料中实现关系映射的所有关系； $op$  表示构建的本体中的关系数量； $ai$  表示语料数据与领域知识本体进行匹配时，可表示为本体中类实例数量； $ci$  表示语料数据中所有的实例数量。

根据本体评分和专家意见，针对个别颗粒化过于明显且在概念或功能上具有相似性的类进行了合并；针对映射失败的实体，对本体的层级结构及细分类进行微调，例如为解决大量“值类型”，如设计尺寸范围、材料性能参数数据、公式等在本体中归类困难的现象，本体类增加了“值类型”的属性参数；根据数据中语义关系在概念架构中的分布状态，

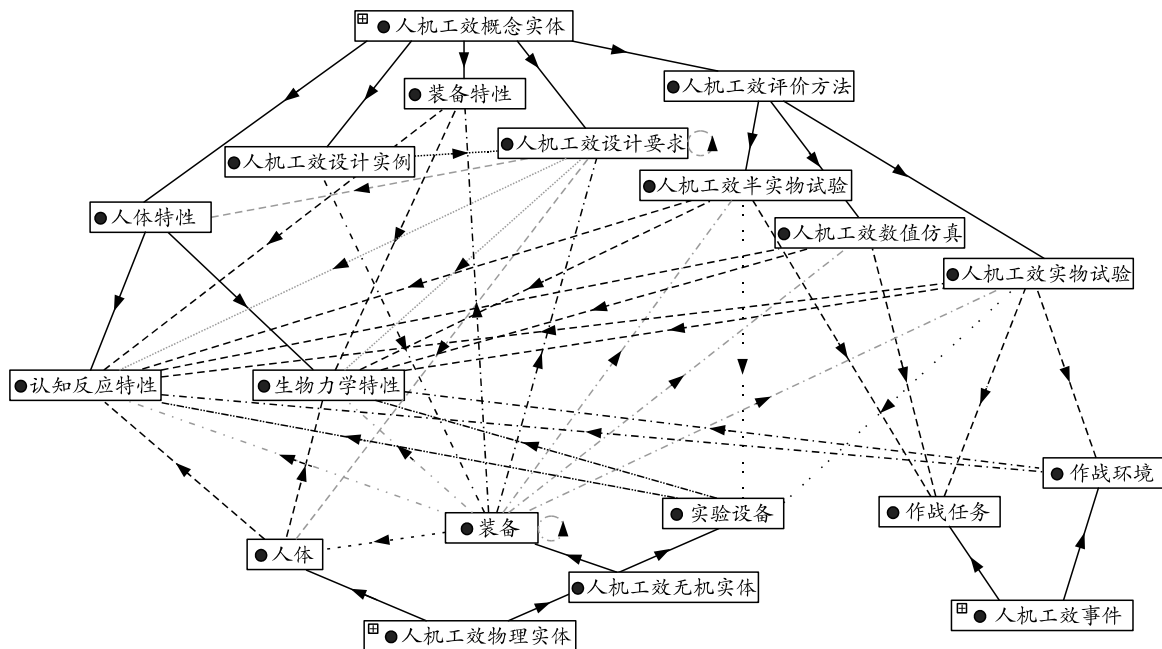


图 2 本体关系网络

## 2 领域非结构化知识联合抽取

知识图谱构建需要对领域内的各种知识来源进行知识抽取，以获取“Subject-Relation-Object”三元组数据集。由于目前单兵人机工效设计与评估领域内的知识大部分为非结构化的文本形式，笔者采用“实体-关系”联合抽取的方式对数据源中的三元组信息进行抽取。

### 2.1 联合抽取模型架构

受领域数据集的规模和深度神经网络训练任务特点的限制，从零开始对领域数据集进行训练，往往难以得到可用的任务模型。BERT (bidirectional encoder representation from transformer) 模型架构为基于 Transformer 编码器的深度神经网络模型<sup>[24]</sup>，

对单兵装备人机工效关系数据进行人工聚类，将其归纳整理到上层的概念架构中。

表 3 为多次迭代改进前后领域本体的各项评价结果，通过改进本体，各项评价结果均有提高，总体评分增加了 0.579。最终形成的本体网络如图 2 所示。

表 3 迭代中的本体质量评分

迭代次数	参数(权重)				综合评分
	CMM (0.4)	PCM (0.4)	PEM (0.1)	IEM (0.1)	
初步构建	0.494	0.947	--	--	0.346
第 1 次迭代	0.740	0.850	0.57	0.75	0.768
第 2 次迭代	0.890	0.904	0.56	0.88	0.862
第 3 次迭代	0.940	0.940	0.52	0.92	0.896
第 4 次迭代	0.995	0.930	0.60	0.95	0.925

可通过 self-attention 机制实现对模型本身的训练。考虑到单兵人机工效领域数据集的规模相对较小，笔者采用经过预训练的 Transformer 模型 ROBERTA<sup>[25]</sup> 作为本文模型的输入解码器 (Encoder)，模型通过通用语料数据集上的预训练后，具有较强的处理自然语言的能力。

笔者采用“ROBERTA + Lambda + Dense”架构的联合信息抽取模型，模型架构如图 3 所示。其中 ROBERTA 模型及其后续部分层为共享层，用于对本文中张量化的编码序列标注，其余部分用于处理张量，以实现 Subject、Object 和 Relation 的预测。

由于单兵装备人机工效知识片段存在较多的三元组重合现象，笔者采用“半指针-半标注”<sup>[18]</sup>的

方式，首先识别句中的 Subject 并进行输出，提取 Subject 在句中的位置编码信息及句子的编码信息；通过条件归一化 (conditional layer normalization) 对

句子编码进行微调，突出 Subject 在句中的语义关系，将处理后的句子编码传入后续网络。根据先验特征对 Subject 对应的 Object 和 Relation 进行预测。

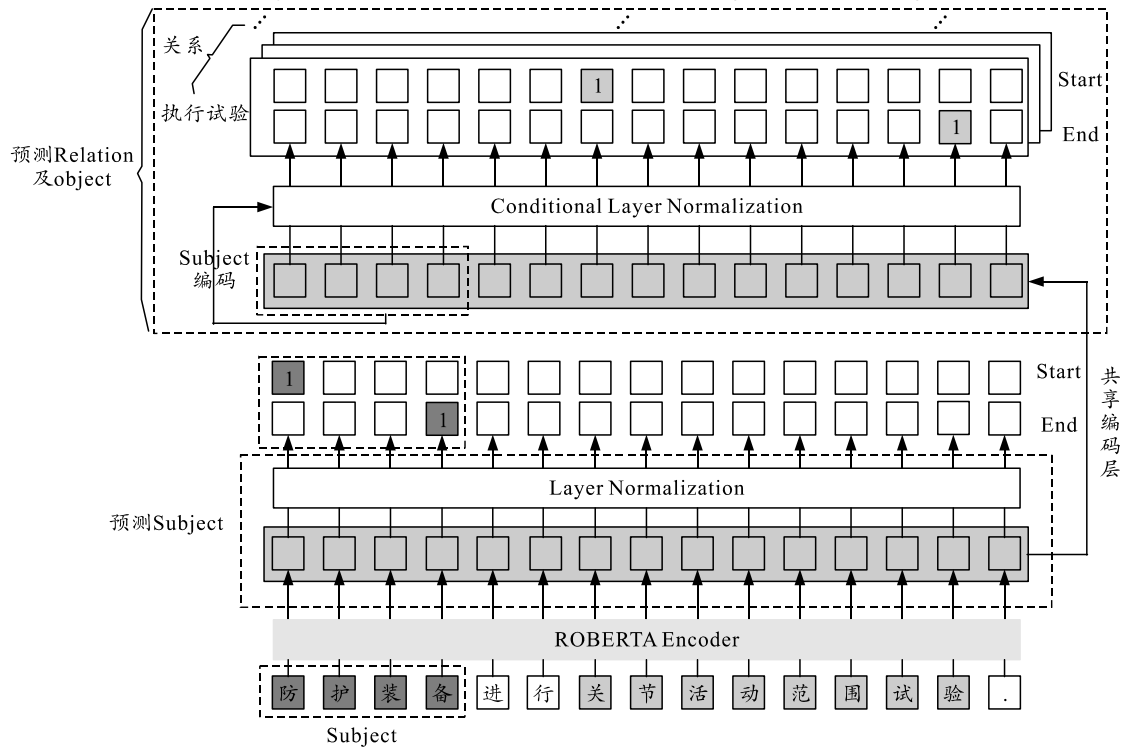


图 3 联合抽取模型架构

### 2.2 模型抽取效果

笔者对联合抽取模型在单兵装备人机工效数据集上的性能及抽取效果进行实验，数据集涵盖《人机工程学》《人机工程与产品设计》《军事装备和设施的人机工程要求》《军事装备和设施的人机工程设计手册》及《单兵装备人机工程建模、仿真与评价》等书籍中的三元组标注数据，共包含 76 种语义关系。实验结果显示，联合抽取模型在单兵装备人机工效数据集上的最高 F1-score 值为 0.82，抽取效果示例如表 2 所示。

表 4 模型抽取效果示例

非结构化文本输入	模型抽取结果
疲劳耐受性试验采用表面肌电测试系统，对比分析了在有负重和无负重 2 种状态下，人体进行战术攀登动作时下肢主要肌肉群疲劳特性	{疲劳耐受性试验, 试验设备, 表面肌电测试系统}, {疲劳耐受性试验, 试验判据, 肌肉群疲劳特性}, {疲劳耐受性试验, 试验任务, 战术攀登动作}, {表面肌电测试系统, 测量特性, 肌肉群疲劳特性}, {下肢, 人体特性, 肌肉群疲劳特性}, {负重, 装备特性影响, 肌肉群疲劳特性}

### 3 人机工效知识图谱应用示例

笔者建立的单兵装备人机工效知识图谱目前主要用于语义搜索，利用数据内在的结构和语义关联，

提供精确和具体的答案，通过实体链接实现对单兵装备人机工效设计与评估知识与文档的混合检索。

由于单兵装备人机工效知识中实体密度大、语义关系复杂，为便于展示，笔者只选取图谱中的部分装备设计要求条目及典型工程案例“携行具背架的设计校核模型”中的数据在 Neo4j 中进行图谱构建和展示。知识图谱局部示例如图 4 所示，示例部分知识的抽取结果中实体的本体分布如表 3 所示。

表 5 联合抽取结果中实体的本体分布

本体	实体个数
装备	47
人机工效实物试验	27
人机工效半实物评价	6
人及工效数值仿真	15
试验设备	12
设计要求	113
人体	97
装备特性	76
人体特性	135
作战环境	16
作战任务	53
人机工效设计实例	8

通过图谱内的实体和关系，可以将单兵装备的设计、仿真、使用和评价等知识实体，以网络的形式关联起来；通过图谱的关联性查询方式，能够快

捷地实现各种知识片段、案例之间的关系查询，实现不同来源知识间的关联推荐。

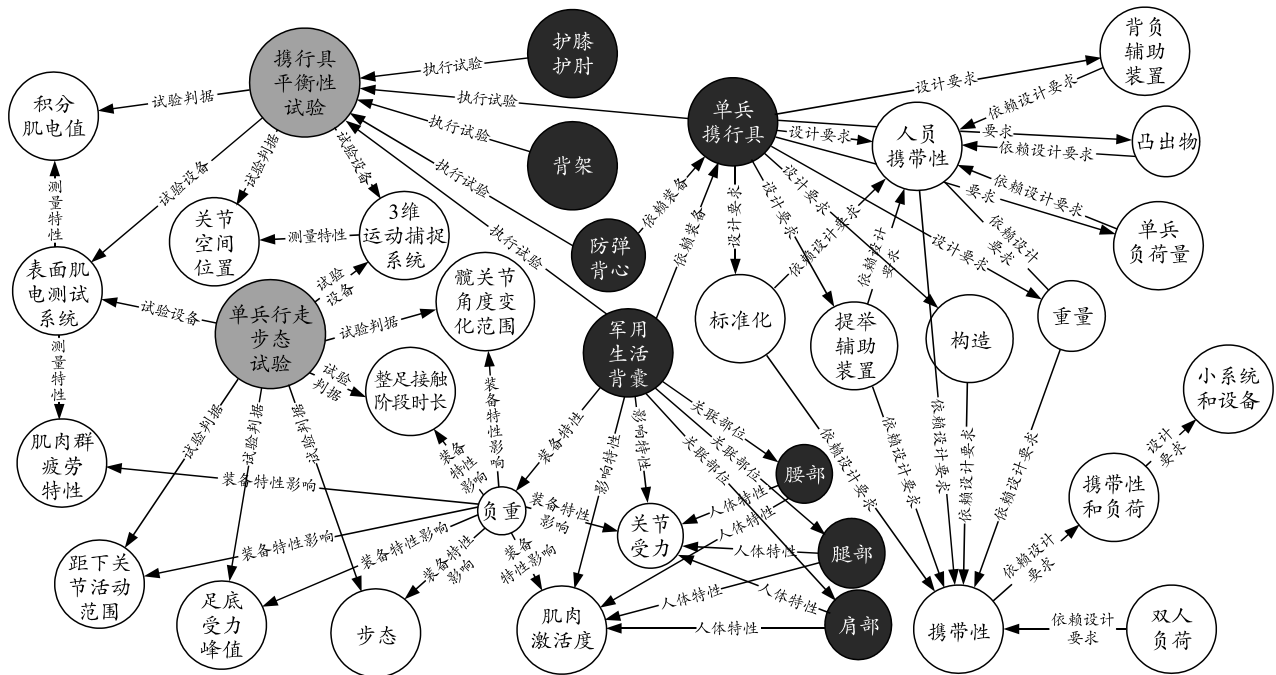


图 4 知识图谱局部示例

具体应用示例如下：

1) 基于装备对其关联人体特性的发散型查询。

通过输入起始装备，对相关人体特性及关联路径进行查询，显示进行装备设计时应考虑的人体特性，在进行试验评价和研究时的重点关注部位。以“携行具分系统”为例，如图 5 所示，能够显示出整个分系统中“军用生活背囊”“作战背包”与“战术背心”与人体的“腰部”“腿部”“肩部”相关。进一步关联出人体的“腰部”应考虑“关节受力”“肌肉激活度”等人体特性。

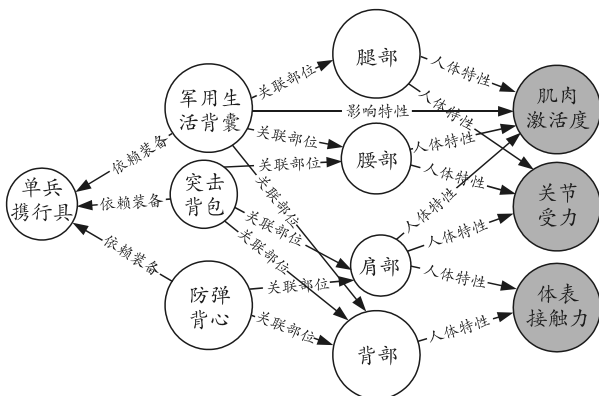


图 5 以装备为中心的关联人体特性查询

2) 装备人机工效设计要求及关联装备查询。

笔者通过基于深度学习的设计要求分类技术，将国家军用标准中的人机工效设计要求项进行了筛选和分类。如图 6 所示，当对“单兵携行具”等装

备的人机工效设计要求进行查询时，知识图谱的输出结果，可获得装备对应的人机工效设计要求项，如“重量”“提举辅助装置”“单兵整体负荷量”“构造”“人员携带性”等。设计人员可根据查询结果对军用标准进行回溯，也可通过“设计要求”类的节点属性值对标准对应的具体文本进行查看。

3) 单兵装备人机工效评价方法查询。

通过对装备对其评价方法和试验查询的形式，辅助研究人员确定试验设备部署方案、选定装备对应评价指标，以便研究人员快速拟定试验计划。如图 7 所示，当查询“防弹背心”人机性评价试验时，得到“佩戴舒适性试验”“平衡性试验”“干涉兼容性试验”等查询结果，关联出试验设备及评价指标，还可关联出采用相同试验类别的其他装备(如军用生活背囊)，提醒设计人员在制定试验方案时是否应统筹考虑。

4 结论

笔者以提高单兵人机工效领域知识的共享与重用为目的，开展了单兵人机工效知识图谱构建相关技术研究，主要结论如下：

1) 提出了基于“一致性检验”与“语料评价”的骨架法应用策略，将单兵装备人机工效领域中的“人-机-环”3 大概念元素与 UMLS 语义框架相结合，基于 protégé 建立了单兵装备人机工效领域知

识本体，对本体网络进行了质量评估与优化。这种骨架法应用策略能够在短时间内构建一套以主客观

评价结合的知识本体，达到对语料数据较为全面的覆盖效果。

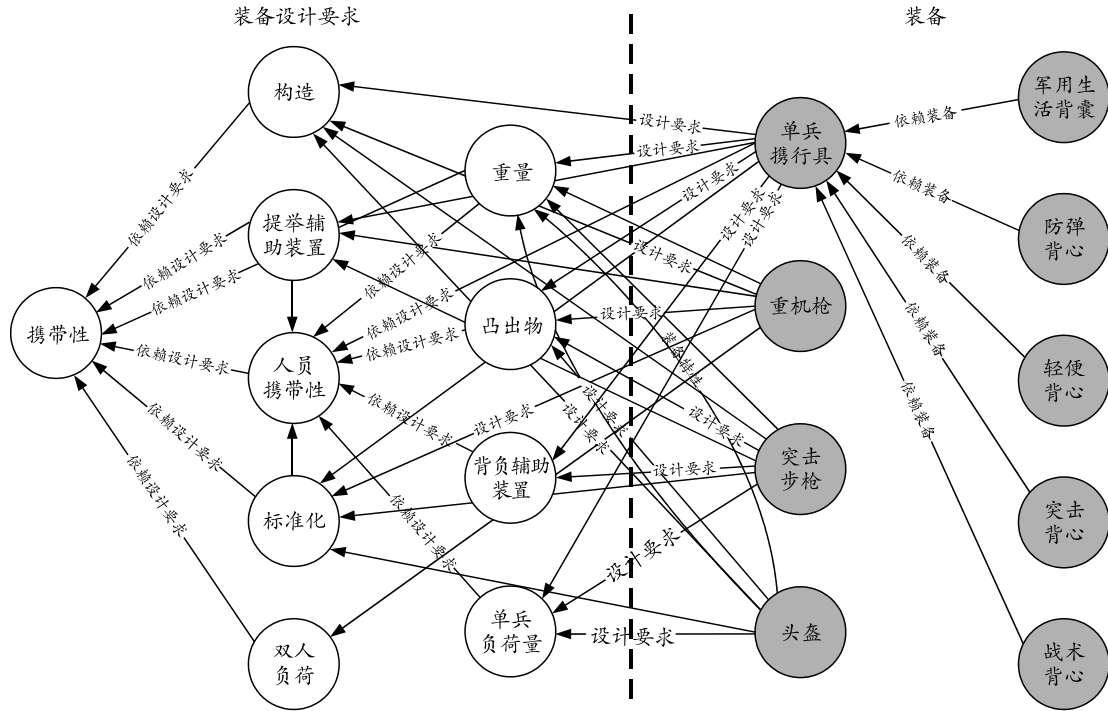


图 6 基于装备的人机工效设计要求查询

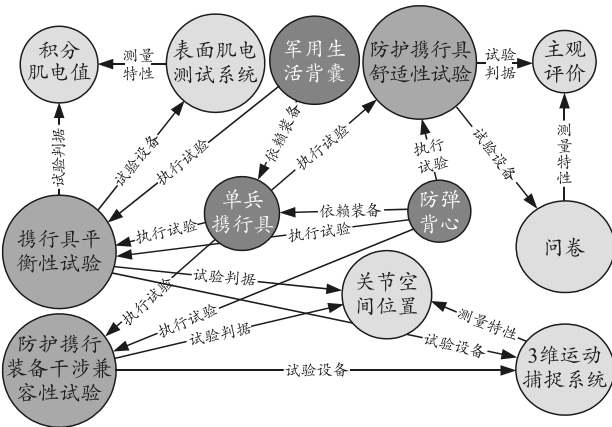


图 7 基于装备的人机工效评价方法关联查询

2) 将基于神经网络的联合抽取模型应用于人机工效领域知识抽取中，实现了对人机工效领域知识中三元组信息的自动化抽取，在重合实体的三元组抽取中具有较高的精度。

3) 对构建的单兵装备人机工效知识图谱应用方式进行了初步讨论，为单兵装备人机工效领域知识在单兵装备设计与评估中的应用研究提供了一种新的技术途径和手段。

由于受限于领域数据集的规模，笔者采用人工为主自动为辅的方式对节点与关系进行本体映射，在后续研究中应进一步优化抽取算法，尝试加入映射模块，以提升大规模领域知识图谱的构建效率；

并对知识图谱融合、知识图谱推理及问答等应用模式开展进一步工作。

参考文献:

[1] 岳丽欣, 刘文云. 国内外领域本体构建方法的比较研究[J]. 情报理论与实践, 2016, 39(8): 119-125.

[2] 王向前, 张宝隆, 李慧宗. 本体研究综述[J]. 情报杂志, 2016, 35(6): 163-170.

[3] 任飞亮, 沈继坤, 孙宾宾, 等. 从文本中构建领域本体技术综述[J]. 计算机学报, 2019, 42(3): 654-676.

[4] 廖莉莉, 沈国华, 黄志球, 等. 本体评估方法研究综述[J]. 计算机应用研究, 2015, 32(3): 647-651.

[5] 赵军, 刘康, 何世柱, 等. 知识图谱[M]. 北京: 高等教育出版社, 2018: 3.

[6] ZELENKO D, AONE C, RICHARDELLA A. Kernel methods for relation extraction[J]. Journal of machine learning research, 2003, 3(Feb): 1083-1106.

[7] ZHOU G D, SU J, ZHANG J, et al. Exploring various knowledge in relation extraction[C]//Proceedings of the 43rd annual meeting of the association for computational linguistics (acl'05). 2005: 427-434.

[8] CHAN Y S, ROTH D. Exploiting syntactico-semantic structures for relation extraction[C]//Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. 2011: 551-560.