多模态知识图谱构建技术及其在军事领域的应用综述

姚 奕,陈朝阳,杜晓明,姚天磊,李青尚,孙鸣蔚
中国人民解放军陆军工程大学 指挥控制工程学院,南京 210007

摘 要:随着数据资源类型的丰富与大模型技术的发展,能够处理多源异构数据的多模态知识图谱(multimodal knowledge graph,MMKG)以出色的数据处理与管理能力而被广泛关注。结合领域需求与特性,对多模态知识图谱构建技术及其在军事领域的应用展开总体概述。基于传统文本知识图谱的相关概念,对多模态知识图谱的基本概念、研究现状进行梳理,分析总结了多模态信息抽取、多模态实体链接、多模态表示学习三个多模态知识图谱构建的关键技术,以及大模型技术在多模态知识图谱构建过程中的应用,探讨了多模态知识图谱在军事领域中的应用场景。最后结合大模型热点和军事需求,对多模态知识图谱构建技术的发展前景及军事应用进行总结与展望。

关键词:多模态知识图谱;构建技术;大语言模型
文献标志码:A **中图分类号:**TP391 doi:10.3778/j.issn.1002-8331.2404-0285

Survey of Multimodal Knowledge Graph Construction Technology and Its Application in Military Field

YAO Yi, CHEN Zhaoyang, DU Xiaoming, YAO Tianlei, LI Qingshang, SUN Mingwei
College of Command & Control Engineering, Army Engineering University of PLA, Nanjing 210007, China

Abstract: With the rich types of data resources and the development of large language model technology, the multimodal knowledge graph (MMKG) that can handle multi-source heterogeneous data has been widely concerned because of its excellent data processing and management capabilities. Combined with the requirements and characteristics of the field, this paper gives a general survey of the construction technology of multimodal knowledge graph and its application in military field. Based on the relevant concepts of traditional text knowledge graph, this paper summarizes the basic concepts and research status of multimodal knowledge graph, analyzes and summarizes the key technologies of multimodal knowledge graph construction, which are multimodal information extraction, multimodal entity link and multimodal representation learning, and the application of large language model technology in the process of multimodal knowledge graph construction, discusses the application scenarios of multimodal knowledge graph in military field. Finally, combined with the hot topics of large language model and military requirements, the development prospect and military application of multimodal knowledge graph construction technology are summarized.

Key words: multimodal knowledge graph; construction technique; large language model

19世纪,德国理学家赫尔姆霍茨在生物学领域提出了模态(modality)这一概念,特指生物凭借感知器官与经验来接收信息的通道。在工程领域,模态是指在雷达、红外线、加速度计、电磁波等多种通道内传输的文字、图片、语音、视频等信息类型。多模态是指将多种模态数据在计算机内进行融合交流、协同作用的状态[1]。2012年,Google公司提出知识图谱(knowledge graph,KG)的概念[2]。知识图谱旨在描述客观世界的概念、实体、事件及其之间的关系[3],本质上它是以实体、属性等为节点,以实体、属性之间的语义关系为边而形成的语义网络图,实体、关系、属性在早期知识图谱中都是以文本形式存在。随着海量多源异构数据的剧增以及传播媒介中信息载体的变化,视觉、听觉等模态信息逐渐取代文本模态信息成为信息传播的主要载体,社会各领域

基金项目:国家自然科学基金(62273356,61806221);高层次科技创新人才自主科研项目(KYZYJKKC0024001)。
作者简介:姚奕(1981—),男,博士,教授,研究方向为智能指挥控制、知识图谱;陈朝阳(2000—),通信作者,男,硕士研究生,研究方向为多模态知识图谱,E-mail:18654392389@163.com;杜晓明(1970—),男,博士,教授,研究方向为指控理论与工程、知识图谱;姚天磊(2000—),男,硕士研究生,研究方向为深度强化学习;李青尚(2002—),男,硕士研究生,研究方向为多模态知识图谱;孙鸣蔚(2002—),女,硕士研究生,研究方向为多模态知识图谱。
收稿日期:2024-04-18 **修回日期:**2024-07-11 **文章编号:**1002-8331(2024)22-0018-20

对知识图谱的需求已经不再满足于单一的文本符号表示，开始提出将文本、视觉、听觉等不同模态的信息融合在一起来增强知识表达的需求，但由于多模态数据之间存在异构性，多模态数据并没有得到很好地处理和利用。

在2022年OpenAI推出ChatGPT自然语言处理模型之后，OpenAI相继推出了GPT-4 Turbo多模态数据处理模型、Sora文本转视频模型[4]、Voice Engine语音生成模型等多个包含大语言模型(large language model，LLM)和多模态大语言模型(multimodal large language model，MLLM)的生成式大模型，实现了从自然语言处理向多模态数据处理的巨大跨越，但因生成内容缺乏事实真实性和可靠性，使得大模型面临着人工智能幻觉、专业领域实用性差等重大挑战。知识图谱具备准确、专业的知识库，对事实性、专业性的知识处理具有非常高的准确性和可靠性，可以用来弥补大模型存在的弊端，但当前知识图谱专注于处理单一模态数据，缺乏对多模态数据的处理能力，因而多模态数据之间的隐含关系并没有得到有效利用，无法有效完成信息融合、推理等工作。

随着阿里巴巴集团新零售多模态知识图谱AliMe MKG[5]、M^{2}ConceptBase[6]等多模态知识图谱的开发与应用，多模态知识图谱因拥有类型和数量更为全面、准确的知识库和强大的多模态数据处理能力而被广泛研究，其能够消除多模态数据之间存在的异构性，并根据不同模态的数据关联挖掘隐含信息，实现知识融合、推理等热门应用。因此，在知识图谱的基础上，研究者们展开了对多模态知识图谱的构建和应用的研究。Zhu等人[7]综述了多模态知识图谱的构建工作、在处理特定问题时的优势、在技术与实际层面的应用以及发展前景和挑战，分析了多模态知识图谱结构和应用中不同解决方案的优缺点。Peng等人[8]综述了现有多模态知识图谱构建原理的优缺点，创新性地提出将实体重命名为节点，重命名后节点一词包含实体、属性和概念三个范围，使得指代更加准确，但该综述更多的是举例论证一些理论，缺乏对关键构建技术的总结分析。Chen等人[9]对多模态知识图谱的发展、构建技术以及在推荐系统、生物医学等方面的应用实例进行了全面的综述，着重讨论了多模态知识图谱构建的关键技术。陈烨等人[10]综述了多模态知识图谱的构建方法和技术，以及在推荐系统、人机交互等方面的应用，归纳了基于属性和基于实体的两种构建方法的主要思路，但缺乏对现有构建技术的对比分析。陈佳云等人[11]首次对多模态知识图谱在农业领域的研究展开综述，并对农业多模态知识图谱在农业智能问答、病虫害识别等方面的应用展开详细介绍，但对于多模态知识图谱的构建技术并没有进行深入调研与分析。

军队信息化与智能化不断发展，军事领域涌现出大量以地理信息、目标定位为代表的结构化数据，和以视频、图像、音频以及文本为代表的非结构化数据，甚至还包括人类指挥人员和参谋人员的指挥艺术、作战风格等隐性认知知识，这些数据呈现要素维度多、来源范围广、术语专业性强、更新迭代慢、可移植性与交互性差、欺诈性等特点，以人工为主的数据处理模式和单一文本模态知识图谱难以有效应对，满足不了军队信息化、智能化发展的需求。随着深度学习和大模型技术的发展与成熟，涉及多模态数据处理的信息获取、信息融合、推理预测等技术得到创新突破，军事数据呈现的弊端逐步被解决，多模态知识图谱逐渐成为应对军事领域存在的多重挑战的重要途径，并取得了一定的成效。

通过上文分析发现，现有多模态知识图谱类综述缺乏对大模型技术在多模态知识图谱构建过程中的运用进行梳理与总结，并且文章聚焦于通用领域的应用，少有文章系统梳理多模态知识图谱在军事领域的应用实例、挑战以及发展前景等。因此，本文拟从技术和应用视角出发，在现有多模态知识图谱构建技术的基础上，创新地综述大模型技术在多模态知识图谱构建过程中的运用以及多模态知识图谱在军事各领域的应用实例和发展前景。

1　多模态知识图谱

2019年，多模态知识图谱概念被提出[12]，它是指在以文本为实体的传统知识图谱基础上，增加了视觉、听觉、触觉、信号等多种模态实体数据与不同模态实体数据之间的语义关系，使得实体的数据类型不再是单独的文本，而是文本、视觉、听觉、触觉等多种模态，如图1所示。多模态知识图谱因其数据类型复杂，从而具有较为庞大的数据量。根据多模态知识图谱表示方式的不同，多模态知识图谱又可以分为基于属性表示的多模态知识图谱A-MMKG(MMKG with multimodal data as attribute values)和基于实体表示的多模态知识图谱N-MMKG(MMKG with multimodal data as entities)[7]。A-MMKG可以理解为存在表达相同含义的两个不同模态信息，一种模态表现为另一种模态的属性信息；N-MMKG可以理解为存在表达相同含义的不同模态数据表现为不同的实体，通过关系相链接。部分多模态知识图谱与数据资源分类如表1所示。

对多模态知识图谱的定义，业界也有不同理解。陈烨等人[10]认为，多模态知识图谱是指基于传统知识图谱，构建多种模态的实体与多模态实体之间的语义关系。Zhu等人[7]认为，多模态知识图谱可以被看作具有多模态化的实体和属性的知识图谱，该知识图谱中的知识符号是多模态化的，与知识符号对应的数据项形式可以是文本、图像、视频等。李华昱等人[21]认为，多模态知识图谱是将多模态信息引入到知识图谱的一种技术，它

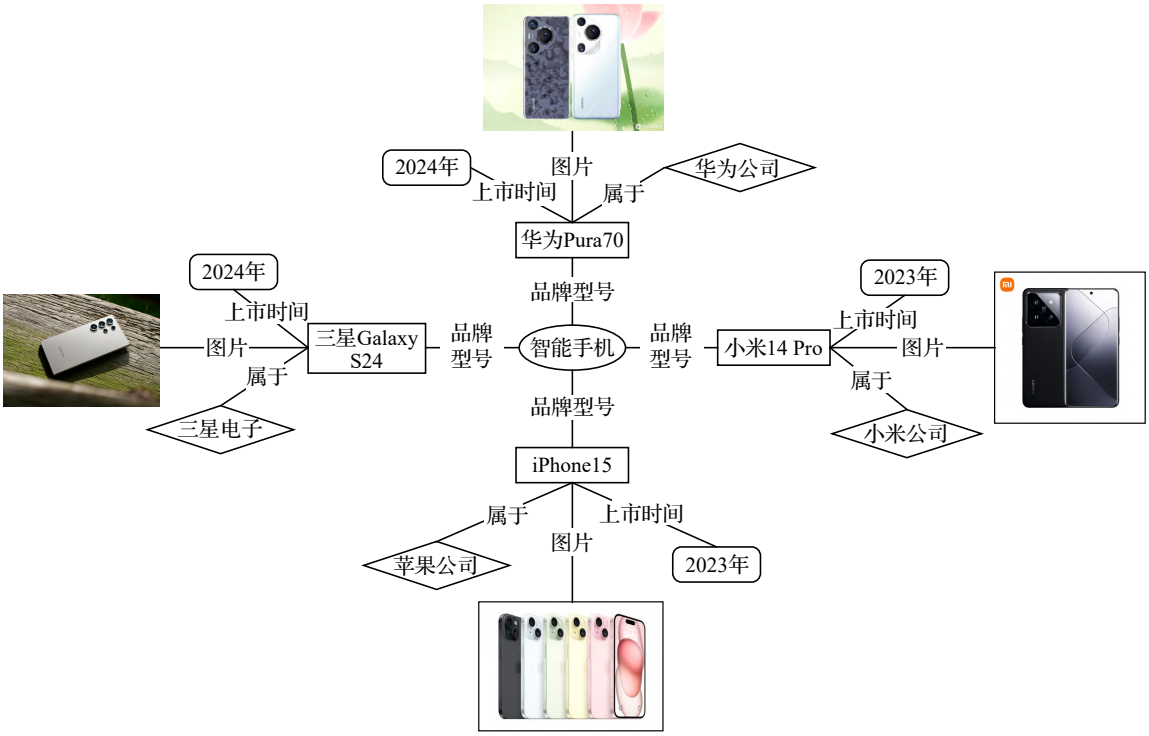

图1　多模态知识图谱形式示例

Fig.1　Example of multimodal knowledge graph form

表1　多模态知识图谱与数据资源

Table 1　Multimodal knowledge graph and data resources

类别	图谱名称	发布时间	数据规模	特点
A-MMKG	ImageNet[13]	2014年	14 197 122张图片,21 841组同义词集合	被广泛认为是2010年的深度学习革命的开始
	DBpedia[14]	2015年	458万个实体和30亿条信息	可以自动与维基百科保持同步更新
	MMpedia[15]	2023年	2 661 941个实体,19 489 074张图像	拥有最大图像集合的多模态知识图谱
	AspectMMKG[16]	2023年	2 380个实体,18 139个实体方面和645 383个与方面相关的图像	第一个考虑实体多方面性质的多模态知识图谱,可以从多个角度去理解实体
	M^2ConceptBase[4]	2023年	951 089张图像,151 776个概念	第一个以概念为中心的多模态知识图谱,每个概念与相关图像和详细文本相关联
N-MMKG	CN-DBpedia[17]	2015年	超过900万个百科实体和6 700万条三元组关系	国内最早推出的也是目前最大规模的开放百科中文知识图谱
	IMGpedia[18]	2017年	1 500万幅图像,4.5亿条视觉相似关系	大型链接数据集,链接数据和多模态数据的结合,图片之间有相似性链接
	ImageGraph[19]	2017年	1 330条关系,14 870个实体,829 931张图像,564 010个三元组	用于探索新的机器学习方法来解决网络提取知识图中的视觉关系查询
	Richpedia[12]	2020年	280万个实体,2 883 162张图像,1.7亿个三元组	利用了图像的配文来识别图像中的其他实体
	MarKG[20]	2022年	11 292个实体,192条关系,76 424个图像,34 420个三元组	第一个用于多模态类比推理的多模态知识图谱

构建了跨模态的实体以及语义关系。Zheng等人[22]认为目前互联网数据呈现多模态特征,因此提出由结构数据(关系三元组)和多模态辅助数据(图像、文本、视频等)组成的多模态知识图谱。多模态知识图谱与传统知识图谱最本质的区别在于实体类型的不同,传统知识图谱主要集中研究文本数据的实体和关系,而多模态知识图谱则是基于传统知识图谱,添加多种不同模态的实体和实体间的语义关系,形成一张具有多种模态信息的知识网络。

2　多模态知识图谱的构建技术

对于多模态知识图谱的构建,当前的构建方式通常分为两种:一是构建单模态知识图谱,在已有文本实体的基础上,通过多模态知识抽取、多模态实体链接、多模态表示学习等技术增加图片、音频等实体类型,扩充知识图谱的实体数量和类型,完成多模态知识图谱的构建工作;二是分别构建以文本作为实体类型的单模态知识图谱和以图片、音频等为实体类型的单模态知识图谱,再利用多模态实体链接、多模态实体对齐、多模态表示

学习等技术将多个类型的知识图谱进行融合,最终形成一个多模态知识图谱。传统知识图谱构建与多模态知识图谱构建的主要差别在于以下几个方面:

(1)模型选取不同。传统知识图谱的构建使用基于文本的自然语言处理技术,通常利用双向长短期记忆网络(bidirectional long short-term memory,BiLSTM)[23]、BERT(bidirectional encoder representations from transformers)[24]等模型处理文本数据。而多模态知识图谱的构建需要采用诸如图像处理、音频处理、视频处理等多模态数据处理技术。同时,多模态知识图谱需要结合深度学习技术对多模态数据进行建模和表示,通过LeNet-5[25]、AlexNet[26]等卷积神经网络模型,SegNet[27]、Mask R-CNN(mask region-based convolutional neural network)[28]和DeepLab[29]等图像分割模型,E-PANN(evolved-plastic artificial neural network)[30]、LAS(listen,attend and spell)[31]等音频处理模型的联合使用来处理多模态数据。

(2)构建复杂度不同。在构建过程中,单模态知识图谱通常只需要考虑一个模态数据之间的关联与交互,不需要考虑不同模态数据之间语义和关系的处理;多模态知识图谱需要考虑多个模态数据之间的关联与交互,挖掘不同模态数据之间潜在的语义关联,判别不同模态数据之间的语义相关性。因此,多模态知识图谱的构建需要对数据进行更加综合与深入的分析与处理。

(3)数据源不同。在知识抽取过程中,单模态的传统知识图谱主要从单一模态的结构化和半结构化文本数据中抽取和分析所需实体与关系,数据源局限于单一模态的数据,易出现语义挖掘不充分等问题;多模态知识图谱可以从结构化、半结构化和非结构化的数据中抽取实体和关系,利用不同模态的语义交互性,得到准确、丰富知识表达。

多模态知识图谱的构建包含多模态知识抽取、多模态实体链接、多模态表示学习等关键技术。多模态知识图谱体系架构如图2所示。在图谱构建过程中,数据的抽取、外部数据实体与图谱的链接以及利用表示学习对数据的处理方式与传统文本知识图谱存在着一定的差异,因此综述重点分析多模态知识抽取、多模态实体链接以及多模态表示学习三种关键技术。

2.1　多模态知识抽取

多模态知识抽取是多模态知识图谱构建的关键技术之一。多模态知识抽取是指从结构化、半结构化、非结构化的多模态数据中进行命名实体识别和关系抽取的过程,多模态命名实体识别的研究旨在从文本、图像、音频等多模态数据中识别并抽取不同数据类型的实体,作为额外的输入来扩展传统基于文本的命名实体识别,为多模态实体链接提供丰富、准确的数据来源。多模态关系抽取的研究旨在从语言序列、相关图像等模态数据

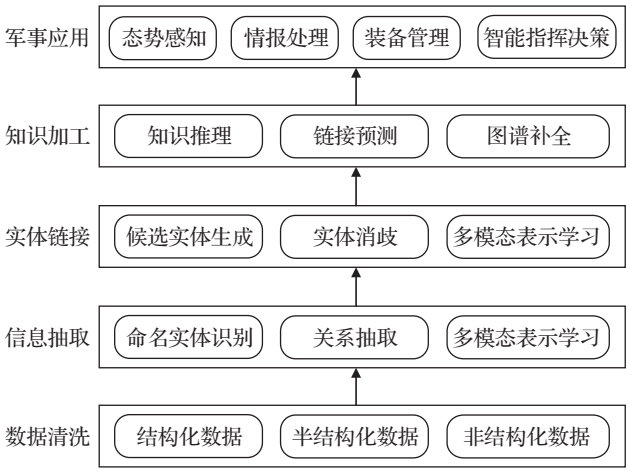

图2　多模态知识图谱体系架构

Fig.2　Multimodal knowledge graph construction framework

中预测三元组头尾实体之间的关系[32],构成一个完整的三元组,更全面、准确地理解多模态数据。根据数据的类型,多模态知识抽取可以分为图像知识抽取、语音知识抽取、视频知识提取等。本节将重点介绍图像知识抽取、语音知识抽取两种常见类型的知识抽取方法,以及新兴的基于大模型的知识抽取方法。

2.1.1　图像知识抽取

在文本知识抽取任务中,通常面临文本信息不充分、歧义大等问题,仅利用文本难以充分理解句子的语义,无法从中抽取精准的实体与关系。随着当今社会新闻和信息传播媒介的进步,图像在各个领域占据越来越重要的地位,在这种情况下,研究者们引入了图像模态信息来补充文本模态缺失的上下文信息和帮助消除文本信息存在的歧义,以完成多模态知识图谱的构建工作。按照抽取模型的类型,可以将图像知识抽取分为基于特征的方法、基于统计的方法、基于深度学习的方法三类。

在早期的图像知识提取过程中,常用尺度不变特征变换(scale-invariant feature transform,SIFT)[33]、方向梯度直方图(histogram of oriented gradient,HOG)[34]、SURF(speeded-up robust features)特征提取算法[35]、ORB(oriented FAST and rotated BRIEF)特征提取算法[36]等基于特征的抽取方法来进行图像信息抽取。基于特征的方法主要通过对图像实体的边缘特征、颜色特征、纹理特征、点特征的检测与分析来抽取图像实体所表达的信息。基于特征的抽取方法过度依赖于人工,只能捕捉图像中的局部信息,对于全局和上下文的信息理解能力有限,环境因素的变化对准确率具有较大的影响且难以处理大规模数据,因此逐渐被基于深度学习的抽取方法所取代。Wang等人[37]提出了一种基于SURF算法和ORB算法的图像匹配算法L-SURB算法,有效解决了ORB算法对图像亮度敏感的问题,其识别速度比SURF算法提高了81.5%。Zhang等人[38]提出了基于SURF算法和改进MIC(minimum intensity change)算法的图像

快速匹配算法MIC-SURF,在保证特征提取精度的前提下,有效地提高了提取速度。部分基于特征的图像信息抽取方法如表2所示。

基于统计的图像知识抽取方法是一种基于机器学习和统计模型的信息抽取方法,通常采用有监督学习算法进行训练和学习,通过从标注好的训练数据中学习统计模型,利用这些训练好的模型对图像像素值的统计特性进行分析和建模,抽取实体与关系。常见的基于统计的算法包括隐马尔可夫模型(hidden Markov model,HMM)、条件随机场模型(conditional random field,CRF)、最大熵模型(maximum entropy,MaxEnt)、支持向量机模型(support vector machine,SVM)等。基于统计的方法可以从大量数据的全局或整体角度自动抽取图像实体的特征,无需人工干预,并且可以通过改变参数或模型来适应不同的任务和需求,具有较强的适应性。但依赖大量的标注数据进行训练和学习,对计算资源提出了更高的需求,且在特定任务中容易出现过拟合等问题,因此逐渐被基于深度学习的抽取方法所取代。Zhang等人[39]提出了一种统一的多模态图融合方法。该方法首先用一个统一的多模态图来表示输入的句子和图像,然后堆叠多个基于图的多模态融合层,最后使用条件随机场解码器执行实体识别。Zhang等人[40]提出了一个命名实体识别模型CBCFuFiC。该模型由自适应共同注意力网络(attention calibration network,ACN)、升级版BiLSTM-CRF模型、门控多模态融合模块和过滤门模块组成,能够很好地处理文本和视觉信息,解决多模态数据带来的噪声问题,提高命名实体识别的准确率。Lu等人[41]提出了基于注意力的命名实体识别模型。该模型利用视觉注意力模型、BiLSTM-CRF序列标注模型和门控机制模块识别最相关实体,消除噪声,提高命名实体识别准确率。基于统计的图像信息抽取方法如表3所示。

随着图形处理器单元的巨大进步和多模态数据的剧增,基于深度学习的抽取方法迅速发展起来,显著提高了抽取任务的准确性和效率[46]。当前研究中,通过循环神经网络(recurrent neural network,RNN)、卷积神经网络(convolutional neural network,CNN)、图神经网络(graph neural network,GNN)、Transformer模型等的使用和不同神经网络模型的结合使用,以及多模态表示学习技术的融入,抽取图像信息所表达出的丰富、准确的语义知识,可以为多模态知识图谱的构建提供广泛且真实的数据来源。各种深度神经网络模型的出现和使用使得图像信息抽取步入了一个新的阶段。Xu等人[47]提出了一个基于流水线的多模态军事信息抽取框架,采用BERT模型和YOLO(you only look once)模型处理文本和图像信息,在CCKS-2022竞赛中获得了第一名。Wang等人[48]提出了一个基于Transformer模型的具有精炼跨模态注意力的框架CAT-MNER。该框架构建了一个R-Transformer模型,使用从知识库中扩展的一些实体标签词来修改跨模态注意力,完成对文本-图像的信息抽取。Sun等人[49]提出了一种基于文本-图像关系推理的命名实体识别模型RpBERT。该模型使用ResNet从图像中提取视觉特征,使用BERT模型处理文本信息,将处理后的编码信息输入到RpBERT模型中,通过跨模态实体之间的关系推理,充分利用视觉信息完成信息抽取。Zhang等人[50]提出了一种结合视觉对象密度的硬样本挖掘策略和去偏对比学习损失的去偏对比学习方法,以减轻实体数量和实体类型之间存在的偏差。Wang等人[51]提出了一种基于门控分层多模态融合和对比训练的模型(gated hierarchical multimodal fusion and

表2　基于特征的图像知识抽取方法

Table 2　Image knowledge extraction methods based on feature

方法	提出时间	优点	缺点
尺度不变特征变换(SIFT)[33]	1999年	对图像的尺度缩放、旋转、平移、光学等变化具有较好的稳定性,对于局部变化和部分遮挡具有鲁棒性	无法有效处理边缘光滑的图像,实时性差
方向梯度直方图(HOG)[34]	2005年	能够较好地捕捉局部形状特征,对图像的光学、几何形变等变化具有较好的稳定性,特征表示能力强	难以处理遮挡图像,对噪点相当敏感,需要根据场景变化不断调整参数
SURF特征提取算法[35]	2006年	处理速度快,对图像的旋转、尺度、光学等变化具有较好的稳定性,匹配精度高	无法有效处理纹理较弱和边缘平滑的图像,对视角和仿射变换敏感
ORB特征提取算法[36]	2011年	处理速度更快,具有实时性,对图像的旋转、尺度等变化具有较好的稳定性,匹配精度高	无法有效处理纹理较弱和边缘平滑的图像,对光学变化敏感,特征表示能力弱

表3　基于统计的图像知识抽取方法

Table 3　Image knowledge extraction methods based on statistics

方法	提出时间	优点	缺点
最大熵模型[42]	1957年	结构紧凑,通用性好	训练复杂,时间开销大
隐马尔可夫模型[43]	1966年	泛化能力强,灵活性高,对数据观测有较好的适应性	参数估计困难
支持向量机模型[44]	1995年	适用于高维空间,内存利用效率高	难以训练大规模数据集,对参数、核函数选择敏感
条件随机场模型[45]	2001年	可解释性强,模型表现力强,准确率高	特征选择困难,训练时间长

contrastive training,GHMFC),以发现多模态数据之间的细粒度关联,挖掘多模态数据之间的深层次含义,实现细粒度的信息抽取。

为了解决单模型处理图像数据暴露出来的问题,研究人员尝试将具有不同特点的神经网络模型结合使用来进行图像处理,出现了CNN-RNN模型[52]、CNN-Transformer模型[53]、GNN-CNN模型[54]等联合模型。联合模型的出现,充分利用了不同类型神经网络模型的优势,使得图像处理任务能够更加全面地考虑到图像的不同特征和属性,更加充分地利用数据的全局与局部信息,提高了对数据的理解和描述能力。Dhiravidachelvi等人[55]提出了HCNNRNN-AHB模型,有效地识别眼底图像渗出物并准确分类,在公开数据集上达到了97.4%的准确率。Chen等人[56]提出了一种用于从遥感图像中精确提取湖泊信息的CNN-Transformer联合模型LEFormer,在两个公开数据集上分别达到了90.86%和97.42%的准确率,抽取准确率大大提升。Parseh等人[57]提出了一种GNN-CNN联合模型G-CNN,能够端到端地理解图像场并准确识别与抽取信息,在三个公开数据集上分别达到了99.91%、96.01%、85.32%的准确率。部分基于深度学习的图像信息抽取方法如表4所示。

单模态文本关系抽取模型在面对长度偏短、嘈杂或缺乏有效上下文信息的数据时,由于无法理解实体之间存在的依赖关系,造成关系抽取效率和准确率不高。社交媒体中图像数据的出现,补充了缺失的语义信息,有助于精确地抽取关系,利用视觉信息来补充语义关联成为关系抽取研究的热点。多模态关系抽取任务需要充分捕捉图像细粒度实体与文本上下文存在的跨模态依赖关系,并通过推理预测挖掘实体之间潜在的关系,以精准构建三元组,获得准确且丰富的信息表达。Wu等人[63]提出了一个新的关系抽取框架,该模型能够同时实现输入信息筛选和潜在信息开发利用,实现了多模态实体关系之间的深度挖掘。Lu等人[64]提出了一种结合双向门递归单元(bi-directional gated recurrent unit,BiGRU)和多头注意力机制(multi-head attention,MHATT)的军事领域关系抽取方法,提高了军事领域关系抽取的准确率。

在多模态数据中,数据源的多样性提供了丰富的数据资源,但也带来了数据噪声和不完整性的挑战。由于不相关信息的干扰和部分数据的缺失,使得信息抽取模型在数据理解过程中产生错误的判断和预测,影响信息抽取的准确率和效率,尤其是在高风险、高技能领域,错误信息的抽取将会导致不可挽救的后果。为了克服数据噪声和不完整性对多模态信息抽取带来的挑战,Yu等人[65]提出了一个多模态命名实体识别任务GMNER(grounded multimodal named entity recognition)与H-Index模型,旨在根据文本实体精确定位对应视觉对象实体,生成实体类型区域三元组,消除噪声带来的影响。Lee等人[66]提出了一种强健的战术地图融合技术(robust tactical map fusion technology,RTMF),采用基于带有生成对抗网络的条件变分自编码器的多元图像完成网络PICNet,实现对敌人遮挡部分的恢复,提高了军事目标识别与抽取的准确率。余晓晗等人[67]提出了一种基于改进MAE(masked autoencoders)的装甲车辆目标前景遮挡部分补全方法,通过分割图像语义,补全图像前景被遮挡部分,识别补全后的图像,提高军事目标识别与抽取的准确率和可靠性。

相关图像知识抽取方法如表5所示。

2.1.2　语音知识抽取

语音作为一种信息载体,承载了发出者的身份、情感、语种等诸多类型的信息。相较于文本知识中单一的文字表达,语音知识因承载了多种类型的表达信息,所

表4　基于深度学习的图像知识抽取方法

Table 4　Image knowledge extraction methods based on deep learning

方法	提出时间	优点	缺点
AlexNet[26]	2012年	引入ReLU激活函数提高非线性建模能力,引入Dropout正则化抑制过拟合,首次使用GPU加速	对计算资源需求高,容易出现过拟合问题
VGGNet[58]	2014年	较小的卷积核,结构简单,通道数较多,具有深度提取能力,迁移学习性能好	计算复杂度高,需要耗费更多计算资源,参数量较大,容易出现过拟合问题
GoogLeNet[59]	2014年	引入了Inception模块,有助于提取丰富多样的图像特征,层数深,参数量少,计算效率高,引入了辅助分类器,克服了梯度消失问题	需要规模大的训练数据集,对计算资源需求高,对小尺寸目标的检测较差
ResNet[60]	2015年	具有强大的特征表示能力,网络参数共享,引入残差学习和跳跃连接,解决了梯度消失和模拟退化问题,泛化能力强	需要大量的计算资源来训练和推理,存在特征失真问题,深度残差网络中有大量的冗余,感受野不够大导致有效深度不够
CNN-RNN模型[52]	2016年	可以利用上下文信息来实现对图像信息的精确抽取,可以处理可变长度的序列	无法实现细粒度的精准识别与预测,模型复杂,计算资源消耗大,需要大量标注数据
MobileNet[61]	2017年	轻量化模型,使用深度可分离卷积,使得计算量小,参数少,推理时间短,泛化能力强	难以捕捉全局信息,部分卷积核容易被训练废掉
EfficientNet[62]	2019年	参数量小,计算效率高,可以捕获图像的深层次特征,	处理较大尺寸的图像时计算复杂度较高

表5　图像知识抽取方法

Table 5　Image knowledge extraction methods

分类	代表作品	提出时间	数据集	优点	缺点
基于特征的抽取方法	L-SURB算法[37]	2018年	亮度不同的图像	速度快,能够有效处理不同亮度环境下的特征提取	只能处理静态目标,无法有效处理运动目标
	MIC-SURF算法[38]	2015年	256×256的经典图像	能够处理模糊和噪声较大的图像,具有实时性和适应性	计算过程内存消耗大
基于统计的抽取方法	UMGF方法[39]	2021年	Twitter2015、Twitter2017	能够充分利用细粒度语义之间的交互,不需要考虑不同模态语义单元之间的对应关系	MLP容易造成过拟合问题,计算资源和数据量需求大,训练时间较长
	CBCFuFiC模型[40]	2018年	Twitter's API[2]	能够精确定位与文本实体高度相关的区域,具有良好的鲁棒性	模型由多个组件组成,体量大,结构复杂
	BiLSTM-CRF+ Visual Attention+ Gate模型[41]	2018年	SnapChat Twitter	能够处理可变长度的序列数据,可以并行计算,处理速度快	细粒度实体识别效果不佳,需要大量的数据标注
基于深度学习的抽取方法	多模态军事信息抽取框架[47]	2022年	竞赛提供的标注训练数据集和未标注测试数据集	能够同时处理多种不同类别的对象,能够解决数据稀疏问题	对小目标检测效果不佳
	CAT-MNER[48]	2022年	Twitter2015 Twitter2017	模型体系结构简洁,能够处理训练数据有限的情况,泛化能力强	计算量大,模型本身无法编码位置
	RpBERT[49]	2021年	TRC MNER dataset、MNER dataset of Snap Research	具有较强的全局感知理解能力	需要大量的计算资源,超参数数量多且调整为合适的超参数困难
	去偏对比学习模型[50]	2023年	Twitter2015 Twitter2017	能够有效解决视觉对象与文本对象之间存在的数量、类型不一致问题,实现深层次语义挖掘	对比学习中的正负样本构造困难
	GHMFC模型[51]	2022年	WikiMEL Richpedia-MEL	能够充分挖掘并利用跨模态信息的细粒度关系,对噪声具有良好的鲁棒性	模型采用流水线方法,存在中间过程的错误传播,数据量需求大
	CNNRNN-AHB模型[55]	2023年	E-Ophtha DIARETDB1	能够有效解决噪声干扰,全局搜索能力强	初值和参数敏感,缺乏适应性
	LEFormer模型[56]	2024年	SW dataset QTPL dataset	能够消除数据噪声的影响,实现精确抽取	需要大量的训练数据
	G-CNN模型[57]	2023年	MIT67、SUN397 UIUC Sports、Scene40	能够端到端处理数据,解决场景模糊带来的挑战	不能很好地利用不同场景之间关系以及上下文有关信息
	基于信息筛选,同时利用多模态关系抽取模型[63]	2023年	MRE	过滤噪声数据,根据输入信息生成抽取信息主题,实现潜在信息的开发利用,关系抽取准确率高	抽取性能依赖于场景图解析器的质量和输入信息的质量
	BiGRU-MHATT模型[64]	2021年	SemEval2010-Task8 军事情景语料库	能够处理长文本数据	泛化能力有待提高,无法实现细粒度关系抽取
	GMNER任务、H-Index模型[65]	2023年	Twitter-GMNER	采用端到端方法,直接生成三元组,避免了信息的错误传播	需要大量的数据标注
	RTMF技术[66]	2022年	Cityscapes Battlefield4	能够有效处理作战环境中遮挡物,检测速度快	多目标重叠问题处理效果不佳
	改进MAE模型[67]	2023年	ImageNet-1K、Compcars、COCO	模型具有较好的稳定性,能够处理装甲目标大面积遮挡任务	补全后的成像清晰度有待提高

以能够实现更全面、更准确的理解与表达。随着语音抽取技术的进步,语音知识逐渐成为多模态知识图谱构建的重要数据来源。根据抽取方式不同,可以将语音知识抽取分为流水线抽取方法和端到端联合抽取方法两类。

传统的流水线语音信息抽取方法是通过利用梅尔频率倒谱系数法(Mel frequency cepstral coefficient,MFCC)、隐马尔可夫模型(HMM)、短时傅里叶变换(short-time Fourier transform,STFT)、线性预测系数(linear predictive coding,LPC)、离散小波变换(discrete wavelet transform,DWT)、感知线性预测(perceptual linear predictive,PLP)等传统方法对语音信号进行不同的变换和计算,将语音信号转换为表示语音特征的数值形式,将提取出代表语音特征的参数转录成文本形式,再利用BERT模型、BiLSTM模型等对文本进行实体和

关系的抽取。传统的流水线抽取方法存在一个共同的弊端:这些过程通过多个独立的模型对不同阶段的数据进行处理,语音信息承载的多种类型的表达信息在各阶段并不能够实现交互理解与运用,并且在多阶段流程化的转化过程中,会造成一定概率的误差与错误出现,形成错误积累,导致语音抽取准确率低,造成资源浪费。流水线语音信息抽取方法如表6所示。

为了克服传统的流水线抽取方法带来的信息不能实现交互理解与运用,以及错误积累的弊端,研究者们尝试将多阶段处理模型进行联合运用与优化,形成了端到端联合抽取方法,无需手动提取语音特征,从原始语音信号中同时实现实体与关系的抽取,得到一个或多个三元组,以完成多模态知识图谱的构建工作。这种方法将原始语音信号到三元组的整个过程作为一个端到端的联合模型进行训练和推断,具有简化模型流程、避免错误积累、充分理解数据资源、提高识别性能的优势[74]。Wu等人[75]提出了一种新的语音抽取任务SpeechRE。该任务使用端到端联合抽取的方法进行语音信息抽取,以原始音频为输入,以包含原始音频中出现的实体及其关系的一个或多个三元组为输出,并使用上采样技术和伪标签技术来进行数据增强,以解决数据稀缺问题。Ghannay等人[76]提出了一个端到端的语音实体抽取的方法。该方法通过独特的神经结构直接从语音中提取实体,可以同时对自动语音识别和实体识别进行联合优化,实现联合抽取。Chen等人[77]提出了一个端到端的联合抽取模型EA-ASR(entity-aware automatic speech recognition)。该模型集成了Transformer模型、Conformer模型和BERT模型,用于从汉语语音中对语音信息进行端到端联合抽取,直接得到所需实体和关系。部分端到端联合语音信息抽取方法如表7所示。

2.1.3　基于大模型的知识抽取

随着大模型技术的迅速发展,研究人员尝试将大模型技术与多模态信息抽取结合起来完成多模态信息抽取工作。大模型具有强大的推理能力、生成能力和学习能力,通过在多模态数据上进行预训练,大模型可以学习到跨模态的通用表示和上下文信息,并生成高质量的辅助信息,然后通过在多模态数据上进行微调,进而辅助挖掘到丰富的特征表示和语义关联,得到更准确的抽取结果。Li等人[78]提出一个将大语言模型作为连接桥梁的多模态命名实体识别框架RiVEG,利用LLM生成精炼的辅助信息,并将原始实体转换为由原始实体和实体扩展信息组成的新的实体信息,进行多模态命名实体识别和视觉定位。Chen等人[79]提出通过一个代表中间推理步骤的思维链方法(chain-of-thought,CoT),将大语言模型强大的推理和扩展能力蒸馏到一个小模型中,进而准确地抽取多模态实体与跨模态实体之间的关系。Chen等人[80]提出一个双层视觉知识增强多模态大语言模型LION,通过获取深层次语义信息和研究实体的空间坐标信息,使得模型可以捕获更细粒度的视觉信息,提高视觉信息抽取的准确率。Wang等人[81]提出了一个端到端的通用信息提取框架InstructUIE,该框架为信息抽取任务引入实体对提取任务和实体对关系识别任务,

表6　流水线语音知识抽取方法

Table 6　Pipeline technology of speech knowledge extraction

方法	提出时间	优点	缺点
短时傅里叶变换(STFT)[68]	1946年	可以反映信号的时间和频率两个特性,可以处理非平稳信号,具有时频变换的可逆性	窗宽不能进行自适应调整,时间分辨率和频率分辨率不能同时达优,只能提供局部时频信息
隐马尔可夫模型(HMM)[69]	1966年	可以实现对隐含状态数据的挖掘,可以捕捉数据的时序依赖关系	目标函数和预测函数不匹配,不能很好地描述语音信号的动态特性,无法处理长序列
线性预测系数(LPC)[70]	1967年	运算量小,数据处理速度快,实时性强,计算效率高,具有较好的压缩效果	对噪声信号和非理想信号敏感,无法完全保留原始语音的全部信息
梅尔频率倒谱系数法(MFCC)[71]	1980年	具有良好的频谱分辨能力,更好地区分人声和噪声,能够很好地区分不同人的不同发音特征	依赖于语音信号的能量变化,依赖于语音信号的平稳性
离散小波变换(DWT)[72]	1981年	可以很好地保留原始语音的全部信息,具有能量集中性,能够很好地捕捉信号的重要特征	不适用于非平稳信号的处理,缺乏自适应性,针对不同场景选择合适的小波基函数比较困难
感知线性预测(PLP)[73]	1993年	基于人耳感知特性,能够更准确地捕捉语音信号的重要特征,抗噪性能好	计算复杂度较高,对非线性语音信号处理效果不佳

表7　端到端联合语音知识抽取方法

Table 7　End-to-end joint methods of speech knowledge extraction

代表作品	提出时间	数据集	优点	缺点
SpeechRE任务[75]	2022年	Speech-CoNLL04、Speech-ReTACRED	直接使用语音进行关系抽取,得到三元组,避免了信息丢失和错误传播,模型体量小	训练数据需求量大,长音频数据处理效果不佳,音频中说话者风格和感情等信息利用率低
端到端语音实体抽取方法[76]	2018年	DeepSUN	使用多任务学习方法,解决了训练数据不足的问题,不需要输入与输出序列一一对齐	CTC损失函数用作训练数据,会产生重复、错误、不准确的标签,无法挖掘深层语义关联
EA-ASR模型[77]	2022年	AISHELL-NER	模型通用性强	对于同音字和多音字实体的抽取效果不佳

提高通过大语言模型进行信息提取的准确率。He等人[82]提出一个多模态关系抽取框架EMRE2llm,利用大语言模型提供常识推理和外部知识库功能,以提高多模态关系抽取的准确率。王震宇等人[83]提出了一种基于大语言模型的两阶段多模态关系抽取模型PLFM(prompt-based LLM for MRE),通过LLM生成准确的辅助信息进行关系预测,提升多模态关系抽取的性能。Cui等人[84]提出了一个视觉关系提取模型OpenVik,通过检测到的关系所在的区域提示MLLM,生成与区域知识相关的无格式知识完成视觉关系的抽取。基于大模型的知识抽取方法如表8所示。

2.2 多模态实体链接

如果不考虑多模态信息,传统实体链接方法中外部实体难以链接到正确的多模态知识图谱中的实体,特别是当文本相对较短或缺乏上下文信息时。在传统的实体链接任务中,仅利用文本信息的模型难以充分理解句子的语义,无法通过实体之外的辅助信息消除实体链接存在最大的歧义问题,实体之间的隐含推理信息也更加难以发掘与利用,因此研究者们尝试将多模态信息应用在实体链接任务中,来补充文本模态缺失的上下文信息,以及利用隐含推理信息帮助消除文本实体存在的歧义,从而完成多模态知识图谱的构建工作和丰富知识图谱的表达。根据链接步骤,多模态实体链接任务可以分为候选实体生成和实体消歧两个任务,本节将重点介绍这两个任务以及新兴的基于大模型的实体链接方法。多模态实体链接方法如表9所示。

2.2.1 候选实体生成

候选实体生成是指在给定的多模态知识图谱中生成与检索实体相关的候选实体集,对检索实体与候选实体集中的实体进行相似度计算。候选实体生成的目标是以最小的计算代价在多模态知识图谱中找到最相关的实体,这个过程与信息检索领域密切相关,都专注于从大量数据中找到相关实体。在候选实体生成过程中,利用多模态召回技术中的聚类算法,将多模态知识图谱中与检索数据相似的数据样本划分到同一个聚类中,使得同一聚类内的数据点之间的相似度最大,不同聚类之间的数据点的相似度最小,同一聚类内的实体数据即为候选实体集。Cheng等人[85]提出了一个多模态检索模型IMMR(image-to-multi-modal-retrieval),通过提出的OCLC4R框架将检索与被检索信息分成多个细粒度的类别,采用模态融合和混合并行策略,实现高效检索。Gulzar等人[86]提出了一种有序聚类算法(ordered clustering-based algorithm,OCA),在推荐系统中根据检索对象生成高质量的候选对象,并有效解决推荐系统的用户冷启动和数据稀疏问题。

CNN网络、Siamese网络、Transforme模型、CLIP(contrastive language-image pre-training)模型等深度学习模型可以充分利用多模态数据丰富的语义信息,学习多模态数据之间的匹配关系,计算检索实体与候选实体在编码空间中的相似度或距离,在短时间内从含有海量数据的多模态知识图谱中召回与检索实体相关的实体,并生成候选实体集。Zhou等人[87]设计了一种对比学习方法CLRec与Multi-CLRec,用于提高候选实体集的生成效率,并减少候选实体的生成误差,提高准确率。Xu等人[88]提出了一个zero-shot两阶段候选实体生成模型,通过将超精细的实体类型信息引入到候选实体集生成阶段,学习多模态知识图谱中候选实体的上下文信息,来生成高质量的候选实体集。Luo等人[89]提出了一种多粒度多模态交互网络框架(multi-grained multimodal interaction network,MIMIC),来处理多模态实体链接过程中噪声数据或特定模态的过度影响带来的候选实体集生层准确性差等问题,并捕获视觉实体与文本实体之间相对应的隐含信息。Yang等人[90]提出了一个多提及实体链接的联合学习框架,通过一个联合特征提取模块从视觉和文本的角度来学习上下文和候选实体的表示,然后设计了一个成对训练方案和一个多提及协作排名方法,用于模拟不同提及之间的潜在联系。

表8 基于大模型的知识抽取方法

Table 8 Knowledge extraction method based on large model

代表作品	提出时间	数据集	优点	缺点
RiVEG框架[78]	2024年	Twitter-2015、Twitter-2017、Twitter-GMNER	充分利用大模型技术生成的辅助信息,准确率高	模型体量大,推理速度在一定程度上有所降低
CoT思维链[79]	2023年	Twitter2015、Twitter2017、SNAP、WikiDiverse	充分利用大模型技术的推理和扩展能力,同时缩小了模型的体量	CoT知识的质量受限于母体大模型
LION模型[80]	2024年	OKVQA、IconQA、RefCOCO等12个数据集	能够实现细粒度的视觉理解与抽取,数据需求低,泛化能力强	通用性差
InstructUIE框架[81]	2023年	IE Instructions	泛化和推理能力强,潜在信息挖掘效率高,信息抽取准确率高	微调需要重新训练模型,使得计算资源需求大,训练时间较长
EMRE2llm框架[82]	2023年	MNRE dataset	能够处理数据稀缺问题,具有较强的灵活性	多模型融合,结构复杂,体量大
PLFM模型[83]	2024年	MNRE dataset	能够实现上下文少样本学习	未能考虑非英文数据集的效果
OpenVik模型[84]	2024年	MSCOCO、GSR、VCR	支持细粒度和零样本关系抽取	噪声数据易导致抽取错误

表9　多模态实体链接方法

Table 9　Multimodal entity linking methods

分类	代表作品	提出时间	数据集	优点	缺点
候选实体生成	IMMR模型[85]	2023年	AliProduct(WAB)^2	能够处理大规模数据集,有效消除数据噪声带来的影响	前期处理中检索与被检索信息单独被处理,关联难以发现
	OCA算法[86]	2023年	Amazon dataset	有效解决推荐系统存在的用户冷启动和数据稀疏问题	不能直接处理字符型数据,需要转化数据类型
	CLRec模型 Multi-CLRe模型[87]	2021年	ML-1M、Beauty、Steam等	能够处理大规模数据,计算成本低,已经商用,稳定性强	有效的负样本和参数选择困难
	zero-shot候选实体生成模型[88]	2022年	Ultra-Fine Entity Typing、zero-shot entity linking dataset	泛化能力强,推理能力强,可以解决数据稀缺问题	对于训练数据之外的数据处理效果不佳,跨域知识迁移能力弱
	MIMIC网络[89]	2023年	WikiMEL、RichpediaMEL WikiDiverse	对噪声有良好的鲁棒性,能够有效挖掘多模态隐含语义的相关性	参数选择困难
	MMEL框架[90]	2023年	NYTimes-MEL	同时考虑多个提及实体之间的相关性,可以并行计算,效率高	模型庞大,参数多,计算资源消耗大
实体消歧	联合消歧模型[91]	2021年	M3EL	不需要文本与视觉之间实体对齐	跨领域消歧效果不佳
	DSRM模型[92]	2015年	Wikipedia、AIDA CoNLL2023	相关实体之间的距离在训练期间达到最小化,可以处理未知实体	需要大量人工进行训练数据标注
	zero-shot MNED模型[93]	2018年	SnapCaptionsKB	能够对数据短缺的任务有效消歧	过于依赖上下文信息
	MMGraph模型 SimTri网络[94]	2022年	MMFi、NEEL、DUEL	能够处理未标记数据,可扩展性强,泛化能力强	不规则的文本数据处理效果不佳
	IMN模型[95]	2022年	TweetsMEL Weibo-MEL	利用多模态预训练模型,减少了对标注数据的依赖	需要大量无监督数据进行预训练,无法处理包含多张图片情况
	AMELINK模型[96]	2023年	AMELI	能够自动去除数据集中的噪声,人工干预少,充分利用属性信息	存在错误传播,计算资源消耗大
基于大模型的实体链接	OVEL[97]	2024年	LIVE	能够进行实时和细粒度实体链接	模型依赖于来自小模型的信息,容易造成结果偏差
	LLMEA模型[98]	2024年	DBP_{ZH-EN}、DBP_{JA-EN}、DBP_{FR-EN}	具有较强的鲁棒性和实用性	容易受特殊字符的影响
	GEMEL模型[99]	2023年	WikiDiverse、WikiMEL	实现端到端实体链接,减少了错误传播的概率	仅靠视觉前缀信息,容易造成噪声和图文不匹配问题
	LifeGraph4模型[100]	2024年	Lifelog dataset	充分利用大模型生成的扩展信息	对模糊数据处理效果不佳
	GeMKR模型[101]	2024年	OKVQA-GS112K、OKVQA-WK21M、ReMuq	对数据数量要求低,泛化能力强	缺乏指令微调
	AutoVER模型[102]	2024年	LAION、CC、SBU	能够实现细粒度的视觉定位	处理未标记实体效果不佳

2.2.2　实体消歧

实体消歧通过计算检索的实体与候选实体集中实体之间的相似性得到的相似度,完成候选实体排名,依据排名结果,选择与检索实体最相关的候选实体,完成实体链接工作。首先将多模态实体转换为向量表示,然后将每个模态的特征向量融合成一个综合的多模态特征向量,最后使用深度神经网络将融合后的多模态特征向量映射到一个统一的向量空间中,通过余弦相似度、Jaccard相似度、欧氏距离等方法进行相似度计算,选择相似度得分最高的候选实体作为最终的消歧结果。Gan等人[91]提出了一个联合消歧模型,将多模态数据资源中的文本和视觉实体分别构建两个图谱,形成实体之间的多对多关系,通过使用Gromov Wasserstein Distance方法匹配不同模态实体之间的对应关系,跨域度量多模态实体之间的相似性,实现联合消歧。Huang等人[92]提出了一种基于深度神经网络的深度语义相关度模型(deep semantic relatedness model,DSRM),使用余弦相似度函数度量实体之间的语义相关度。

实体消歧阶段,多模态数据中的实体可能存在多义性或歧义性,即因为一个实体可以在不同语境下表示不同的概念或语义,所以一个实体可以有多个不同的含义或解释,尤其是引入多模态数据之后,丰富了实体表达,但仅靠实体本身无法准确判断实体在当前语境中所代指的具体含义,造成实体指向不明确,需要借助丰富的多模态上下文信息、隐含推理信息和语境信息来帮助理解当前语境中实体的具体含义,获得更全面、更综合的特征表示,明确实体所指,提高实体消歧的准确率。Moon等人[93]提出了一种新的零样本学习(zero-shot)多模态命名实体消歧模型。该模型利用CNN从图像中获得视觉上下文向量,并与从双向LSTM提取的文本上下文结合,有效地解决了对隐藏实体的挖掘工作,获得了丰富的语义信息。Zhou等人[94]提出了用于实体消歧的

MMGraph模型。该模型使用多模态图卷积来聚合视觉和上下文语言信息,以提高命名实体消歧的有效性。Zhang等人[95]创新性地提出了通过多通道迁移学习和元学习在知识层解决多模态实体消歧任务,设计了一个交互式多模态学习网络(interactive multimodal learning network,IMN),以充分理解多模态信息。Yao等人[96]构建了一个大规模数据集AMELI,提出了属性感知的多模态实体消歧模型AMELINK,将属性信息纳入实体消歧过程,辅助实体消歧过程,提高准确率。

2.2.3 基于大模型的实体链接

新兴的大模型因其强大的生成能力和理解能力被广泛应用于多模态实体链接工作中。利用大模型为目标实体生成简要描述和候选实体集,通过大模型强大的推理能力理解跨模态实体的隐含语义和关系,消除歧义,选择最相关实体链接到多模态知识图谱中。Zhao等人[97]首次提出了在线视频实体实时链接任务(online video entity linking,OVEL),并提出一种基于大语言模型作为存储管理器的视频流信息综合管理框架来完成OVEL任务,利用LLM进行候选实体生成和细粒度实体消歧。Yang等人[98]提出了一种大语言模型增强实体对齐模型(large language model-enhanced entity alignment,LLMEA),使用LLM为目标实体生成虚拟等价实体,并利用嵌入在LLM中的隐含语义信息,基于虚拟等价实体与候选实体之间的编辑距离生成候选实体集,最终利用LLM预测最准确的链接实体,完成实体消歧。Shi等人[99]提出了一种基于大语言模型的生成式多模态实体链接模型(generative multimodal entity linking framework,GEMEL),依据多模态上下文信息,使用LLM直接生成目标实体的名称,实现了端到端的多模态实体链接。Rossetto等人[100]提出了LifeGraph4模型,通过使用最大边缘相关性推荐方法和计算大语言模型生成的图像描述编码嵌入与知识图谱中候选实体编码嵌入的余弦相似度,检索选定相关候选实体,生成候选实体集。Long等人[101]提出了一种端到端的生成式多模态知识检索模型(generative framework for multi-modal knowledge retrieval,GeMKR),将视觉特征投影到大语言模型的文本特征空间中,捕捉跨模态的交互,通过训练好的LLM生成与检索实体相关的知识线索,在多模态知识图谱中检索候选实体。Xiao等人[102]提出了一种用于视觉实体检索识别的自回归模型(autoregressive model for visual entity recognition,AutoVER),将对比训练整合到MLLM中,并将检索实体表述为序列到序列的生成问题来改进在多模态知识图谱的大量数据中检索生成候选实体的问题。

2.3 多模态表示学习

在抽取到不同模态数据的特征之后,如何有效地理解和融合这些信息也引起了广泛的关注。通过多模态表示学习,将来自不同模态的数据通过神经网络进行特征抽取并编码,编码后的特征以向量形式映射到一个共享的语义子空间中,进行跨模态对齐、跨模态情感分析、跨模态链接以及跨模态检索等工作。本节将重点介绍基于深度学习的表示学习和基于大模型的表示学习。多模态表示学习方法如表10所示。

2.3.1 基于深度学习的表示学习

在多模态表示学习过程中,利用早期融合、中期融合、晚期融合、跨模态融合、多层次融合等多模态融合技术将不同模态的数据进行交互、融合,提高对多模态数据的理解和处理能力,缩小不同模态之间的异质性差距。Sato等人[103]提出了基于深度多模态表示学习技术的多模态深度学习(deep learning,DL)模型,采用有监督学习和深度多模态表示模型进行训练数据和信息融合,缩小了多模态数据之间存在的异质性差异。Huang等人[104]提出了一种基于多模态表示学习的模型(multi-modal representation learning-based model,MRLM),同时训练全局特征表示学习和多模态特征表示学习两个模块,提高了基于多模态知识图谱推荐的准确率。Yu等人[105]提出了一个基于多模态表示学习的CommerceMM模型,通过结合预训练任务对电子商务主题信息进行多样化和细粒度理解,从而为用户提供更准确的商品推荐。Hu等人[106]提出了一种可扩展深度多模态学习模型(scalable deep multimodal learning,SDML),针对每个模态数据建立特定的网络来将其转换到相同的预定义共享空间,并行训练这些特定的网络,提高了训练效率。

在多模态表示学习中,跨模态翻译方法引起了广泛的关注。跨模态翻译旨在将一种模态的数据转换成另一种模态的数据,并在转化过程中探索跨模态数据之间复杂的语义关联,在一个共享的、跨模态的语义空间中实现不同模态之间的语义对齐,获得更全面、准确的语义表达。Qi等人[107]提出了一种跨模态双向翻译模型(cross-modal bidirectional translation,CBT),有效地捕获图像和文本两个特征空间的跨模态相关性,促进双向翻译。Zhou等人[108]提出了一个跨模态翻译和对齐框架,通过构建两个并行的多模态数据编码器-解码器结构和基于注意力机制的跨模态注意模块,整合模态内信息并生成跨模态表示,探索跨模态数据的相关性,并利用潜在的互补信息。Ye等人[109]提出了一个端到端的跨语音文本网络(cross speech-text network,XSTNet),使用渐进式的多任务学习策略,将语音和文本作为输入,输出转录和翻译的文本,能够充分挖掘不同模态数据之间的隐含关系,探索跨模态数据之间的相关性。

2.3.2 基于大模型的表示学习

随着大量的训练数据和复杂模型结构的引入,大模型技术越来越多地被用于多模态特征提取和编码、多模

表10　多模态表示学习方法

Table 10　Multimodal representation learning methods

分类	代表作品	提出时间	数据集	优点	缺点
基于深度学习的表示学习	多模态DL模型[103]	2022年	包含972个肝结节的自建数据集	数据拟合能力强,可扩展性强	需要大量的训练数据
	MRLM模型[104]	2019年	MovieLens-20M BookCrossing	能够有效提取全局特征,对类别不平衡数据具有较强的鲁棒性,泛化能力强	存在过拟合问题,运算量大,需要更长的时间
	CommerceMM模型[105]	2022年	102 MB的图像文本对和50 MB跨模态交叉对	可扩展性好,具有较强的鲁棒性	对数据敏感性强,计算资源要求高,模型结构复杂
	SDML模型[106]	2019年	PKU XMedia、Wikipedia、NUS-WIDE、MSCOCO	支持多个模态数据的处理,可扩展性强,具有较强的泛化能力,可以并行训练,计算效率高	每个模态对应特定网络,模型体量大,每个模态数据单独处理,难以发现模态间深层次的语义关联
	CBT模型[107]	2018年	Wikipedia、Pascal Sentence、XMediaNet	能够实现细粒度数据利用,泛化能力强	计算资源需求大,需要大量的训练数据
	CMTA框架[108]	2023年	TCGA	数据利用率高	必须对跨模态表示施加对齐约束
	XSTNet网络[109]	2021年	ST datasets MT datasets	端到端翻译,避免了信息错误传播,具有较好的鲁棒性和泛化能力	计算资源需求大,容易出现过拟合问题
基于大模型的表示学习	VisCoT模型[110]	2024年	RefCOCO、RefCOCO+、RefCOg、Visual CoT	可以同时处理标记和非标记以及低分辨率数据,可以实现细粒度数据处理	难以处理包含大量信息的图像和特别复杂的问题
	GraphAdapter框架[111]	2023年	ImageNet等15个数据集	泛化能力强,具有强大的适用性	生成的提示信息过于简单,缺乏多样性
	KAM-COT模型[112]	2024年	ScienceQA	模型体量小,泛化能力强,对训练数据数量要求低	对专业领域的适用性差
	Vengle模型[113]	2024年	GQA、scienceQA等8个数据集	泛化能力强	对数据集的质量和数量要求高
	MCL[114]	2024年	CIRCO、CIRR、MMC、GeneCIS、VQAv2	能够利用大语言模型自动生成三元组,有选择性地进行数据输入	通用性差
	ICL[115]	2024年	ISEKAI ImageNet100	能够实现隐含信息的推理理解	过于依赖已有图像-标签对的准确关联,对长文章处理效果不佳

态知识融合以及推理和预测等表示学习过程中,通过大模型思维链(CoT)[116]、合成学习、共享空间表示等技术,让大模型逐步参与到复杂问题分解成子问题并依次求解的过程,从而有效地融合和表示不同模态的数据,实现复杂问题的推理求解。Shao等人[110]提出了一种VisCoT模型,通过利用视觉思维链来提高多模态大语言模型的推理和可解释能力,实现更准确的数据处理。Li等人[111]提出了一种GraphAdapter框架,通过利用视觉LLM强大的表征能力和下游任务中丰富的类间关系,对文本和视觉空间中的多模态知识进行融合,并与图卷积学习相结合,生成丰富的提示信息,挖掘潜在的关联知识。Mondal等人[112]提出了一种知识增强的多模态思维链推理模型(knowledge augmented multimodal chain-of-thoughts,KAM-COT),在学习过程中借助多模态思维链和LLM生成的知识图谱之外的知识,深度融合多模态信息,实现更深层次的上下文理解。Chawla等人[113]提出了多模态表示学习模型Vengle,利用动态机制将编码后的视觉信息集成到大语言模型中,充分地感知和理解文本和图像之间的关系。Li等人[114]提出了多模态合成学习(multimodal composition learning,MCL),通过利用多模态大语言模型从多模态数据中合成信息和促进视觉特征到语言空间的映射学习,来增强模型对多模态上下文的理解和利用。Tai等人[115]提出了链接上下文学习(in-context learning,ICL),通过多模态大语言模型的因果推理,发现上下文中隐含的因果关联,从而实现多模态大语言模型对未见信息的识别和理解。

3　多模态知识图谱在军事领域的应用

蔡群等人[117]利用海量的文本多源目标情报数据构建了电子目标知识图谱,为电子对抗作战指挥与辅助决策提供了有力支撑。顾丹阳等人[118]提出一种基于本体的知识图谱构建方法,利用知识图谱实现武器装备知识的有效管理和智能检索,以解决武器装备领域数据存在的来源分散、大量冗余、缺乏关联等问题。黄伟春等人[119]提出了利用基于预训练模型和基于规则的军事术语知识图谱构建方法,用于获取和管理军事术语数据中的语义信息。邢萌等人[120]针对军事领域的特殊性提出了军事领域知识图谱构建及应用技术框架,有效推进了指挥信息系统智能化辅助的进程。袁清波等人[121]针对军事指挥控制保障领域的特点,提出了一种融合汉字多特征

的BiLSTM+CRF命名实体识别模型,提升了命名实体识别的准确率。Wang等人[122]利用CBOW(continuous bag-of-words)模型、BiLSTM+CRF模型、BiGRU网络构建了面向目标的杀伤链知识图谱,以完成杀伤链包含的控制设备、传感器设备、打击设备和评估设备四部分的文本数据管理与利用。傅浩等人[123]提出了一种面向军事事件的主题检测与抽取方法,通过使用聚类策略与关键词排名算法完成军事事件的主题检测和主题抽取。成浩等人[124]提出了基于本体的目标情报知识图谱构建方法和基于深度认知的军事目标情报分析方法,有效地提升了军事情报的质量。

通过对上述军事领域知识图谱的构建以及相关技术研究分析可知,当前大多是基于文本模态知识图谱的研究,涵盖情报分析、装备管理、军事术语、指控保障等军事领域的各个方面,少有涉及可以处理图像、语音等多模态数据的多模态知识图谱的研究与分析。现代化战争从机械化向智能化转变,各类信息化和智能化武器装备、传感器以及信息网络技术得到快速发展,多模态军事数据爆炸式增长,以要素维度多、来源范围广、术语专业性强、更新迭代慢、可移植性与交互性差、欺诈性等特点广泛存在于各类军事数据库、移动终端中,高效处理并利用多模态军事数据对于指战员准确掌握战场态势,做出正确决策,形成指挥员信息优势与决策优势将起到重要作用。基于当前知识图谱在军事领域的研究现状,国内外研究者开展了多模态知识图谱在军事领域应用的研究。

3.1 态势感知

现代化战争中,电磁干扰、电子欺骗、认知域作战等新型战术战法的运用使得战机变得稍纵即逝,面对瞬息万变、错综复杂的战场态势,只靠指挥员人脑已经无法快速处理并做出判断,严重影响了指战员对战场态势认知的准确性和时效性[125]。现代化战场态势感知不仅是对当前战场态势信息的感知察觉和认知理解,还是对战场态势变化趋势进行的准确预测和判别,通过利用多模态知识图谱的推理与预测技术,分析敌我双方地理位置信息、敌方装备部署信息、敌方人员关系信息、敌方人员社交媒体数据等海量异构的情报数据,发现潜在的威胁和战争走向等战场数据,形成直观的战场态势,并进行全面、准确的了解和评估,为指挥控制和军事行动提供预警支持。

王昊奋等人[126]为战场态势感知多模态知识图谱的构建,提出了一种结合多种知识表示方法的态势知识统一表示框架,并以表示学习为基础实现数据驱动与知识驱动的融合,同时归纳了适用于面向战场态势感知场景的复合推理方法。黄梓航等人[127]设计了战场环境多模态知识图谱智能服务系统(battlefield environment knowledge graph intelligent service system,BEKG-ISS),提出了基于作战场景认知战场态势的分析流程,提高了智能决策以及战场环境的智能化水平。Lee等人[66]利用知识图谱提出了基于协同智能的实时战场态势感知技术,通过连续学习战场信息的时空变化,以进行战场态势的感知与传达。2012年国外DARPA公布了PLAN X项目,在大规模、实时和动态的网络环境中自动化处理网络地图、作战单元等信息,构建战场网络知识图谱,利用可视化的方式执行战场网络入侵和对抗任务。

3.2 情报处理

军事情报数据来源广泛,结构复杂,存在大量隐含信息,且存在动向性和实时性,如何从海量异构的情报数据中捕获高价值信息并挖掘推理隐藏的战略意图,是情报处理的核心问题。基于多模态知识图谱,构建情报分析与处理体系,根据多模态情报实体间的因果、顺承关系,梳理事件的来龙去脉,挖掘隐藏在表面情报数据背后的情报信息,并对情报真伪进行研判,预测事件的发展和演化进程,准确把握重要事件的走向。

国外将Palantir知识图谱运用在情报分析领域[128],通过与GIS技术的结合以及数据分析与融合的使用,将各单位各领域孤立的海量异构数据,转化为可被利用的指挥决策能力,应用于战场态势分析和预测。在亚丁湾护航等任务中,Palantir知识图谱得到广泛应用。2019年DARPA开展基于时序信息和时间模式的知识导向人工智能推理图谱KAIROS项目,通过人工智能、知识图谱和机器学习技术,发掘并预测多媒体信息与情报信息背后隐藏的战略意图,密切监视世界各地的动态。2021年国内渊亭科技发布了一款分布式战略情报分析平台。该平台通过强大丰富的数据分析模型,构建情报事理信息知识图谱,完成多源多模态情报数据收集、多模态情报数据处理、情报分析可视化、实时情报研判等情报处理工作,极大地提高了情报的价值转化率。

3.3 装备管理

在现代化联合作战的背景下,诸军兵种的各型号武器装备在使用中涌现出海量异构的军事装备数据,合理管理利用这些数据将关系到作战行动的部署与决策。当遭遇敌方导弹、无人机等武器攻击时,通过全球装备多模态知识图谱平台与各类传感器的连接,将捕获到的攻击数据经过数据处理之后链接到多模态知识图谱中,得到攻击武器的各类参数、攻击特点,根据飞行姿态判断出武器的攻击目标,根据攻击武器的型号和状态,在攻击武器的上升段、中段和末端,分别给出最合适的拦截武器、防御与拦截方案,避免重要目标遭受敌方打击。通过多模态知识图谱在装备管理领域的应用,使得指战员能够全方位掌握装备信息,为现役武器装备的使用与维护、新型武器装备的研发提供有力的支撑。

彭京徽等人[129]针对军事装备数据的显著特征,分别构建文本知识图谱和图像知识图谱,通过跨模态实体对

齐,得到了军事装备领域的多模态知识图谱,提高了数据使用效率与管理水平。胡卫等人[130]对军事装备数据知识图谱开展了研究,实现了各模态装备管理数据的分层级多视图的可视化呈现,使得部队装备管理的信息化、智能化水平得到提高。Xu等人[47]在CCKS开源多模态军事装备数据的事件要素抽取竞赛中提出了针对军事装备的多模态信息抽取,为构建军事装备多模态知识图谱提供准确的数据来源,实现了军事装备数据的高效管理与利用。2020年国内渊亭科技发布了自主研发的面向国防领域的超大规模应用级知识图谱平台。该平台具有数据质量高、数据信息完整、数据规模大、可视化操作等特征,贯穿军事作战各个要素、各个环节,基于该平台完成了面向仿真推演领域的装备知识图谱构建、装备科研论证、装备后勤保障等一系列武器装备工作。

3.4　智能指挥决策

基于多模态知识图谱,实现从“多个参谋围着一个指挥员”向“一个精干的反应迅速的智能型参谋”的转变[131]。多模态知识图谱可以整合作战记录、典型战例、训练数据等数据,根据各类传感器以及侦察人员提供的战场态势信息,通过分析战例数据,结合战场态势进行内部推演与任务规划,以文字、图表、语音等形式提出相应的攻防策略,辅助指挥员做出正确的战略决策。

国内宗滕等人[132]根据当前军事数据建设的现状,提出了多模态数据在作战指挥和军事训练等军事领域的智能化应用构想,为实现军队现代化建设提供了有效的数据支撑。李卫星等人[133]提出了一种面向多源数据的新型军事信息系统架构,研究了在现代化联合作战背景下军事领域知识图谱的构建技术,为现代化联合作战中的指挥控制提供了高效、精准、可靠的数据支撑。Lee等人[66]基于知识图谱提出了深度人工智能军事参谋,用于分析管理多智能体感知的战场数据,识别潜在的战场威胁,有力支持指挥员做出正确的决策。2023年国内智谋科技发布了智谋科技知识图谱平台,以可视化分析模式将地理信息和知识图谱相结合,通过对军事行动过程分析与描述,辅助指战员聚焦重点行动,避免冗余信息干扰,实现宏观指挥决策支持。

总体来看,多模态知识图谱在军事领域应用广泛,且潜力巨大。一方面,军事领域需要应对复杂多变的情况,涉及到多维度的空间、时间、属性等信息,多模态知识图谱可以将这些信息进行有效的结构化和组织,提高关键信息的抽取和应用能力,从而实现对海量异构数据的高效利用;另一方面,通过整合文本、图像、音频等多模态数据,构建多维度、跨领域的多模态知识图谱,实现在态势感知、情报处理、装备管理和智能指挥决策领域更深层次的融合与发展,有助于实现对复杂战场环境的全面感知,并为军事指挥决策提供更全面、准确的信息支持。多模态知识图谱的军事应用如表11所示。未来,随着大模型、人工智能、物联网等技术的不断发展,多模态知识图谱在军事领域的应用前景更加广阔,为军事领域的智能化建设和打赢现代化信息战提供重要支撑。

4　展望

大模型,如ChatGPT自然语言处理模型、Sora文本

表11　多模态知识图谱的军事应用

Table 11　Military applications of multimodal knowledge graph

分类	军事图谱名称	提出时间	作用	缺点
态势感知	基于战场态势的知识图谱[126]	2022年	基于知识图谱实现战场态势感知、理解和预测,为作战决策提供有效支撑	响应速度有待提高,错误数据的判别率低,推理能力弱,不能很好地利用部分缺失数据
	战场环境多模态知识图谱智能服务系统[127]	2021年	整合分析战场环境数据,实现战场环境综合情报分析,提供智能辅助决策	功能模块之间交互能力弱
	PLAN X知识图谱	2012年	以简化的作战流程在直观界面上实现与敌方的网络攻击与对抗	作战系统整合不足,未形成高效运行的网络作战系统体系
情报处理	Palantir知识图谱[128]	2004年	通过与GIS技术结合,利用地理信息进行情报融合与分析,指导复杂战场环境下的军事行动	数据库可访问,容易遭到攻击,保密性差
	KAIROS知识图谱[129]	2019年	分析挖掘多媒体信息背后隐藏的线索信息,对情报领域的因果关系进行智能推理	只具备事件模式的理解能力,不具备独立理解并解决问题的能力
装备管理	装备领域多模态知识图谱[130]	2022年	有效地管理和识别军事装备数据,提供准确的战场数据	通过实体对齐融合单模态知识图谱得到多模态知识图谱,跨模态实体间的语义关联难以利用
	军事装备管理数据知识图谱[131]	2022年	整合装备使用管理中的大量数据,应用于装备智能化保障、态势呈现、装备需求论证等领域	图谱去噪能力弱,不能很好地消除噪声对知识图谱带来的影响
	天机·武器装备图谱平台	2020年	以可视化方式保障一系列武器装备工作	平台规模庞大,包含数据量多,需要配备高性能计算设备
指挥控制	军事领域知识图谱[134]	2020年	解决了战略战役级联合作战信息保障面临的挑战,支持军网、民网、数据链等多种传输方式	面向级别高,体量庞大
	智谋科技知识图谱平台	2023年	以可视化模式为指挥决策提供有效的辅助支持	视频知识、多语种知识处理效果不佳

转视频模型、Voice Engine语音生成模型等,由于其新兴能力和高效处理能力,在通用人工智能领域掀起了新的浪潮。随着大模型的出现,人们不禁有一个疑问:在大模型盛行的时代,多模态知识图谱过时了吗?答案是否定的,反而在大模型时代,多模态知识图谱将会得到更好的发展。大模型是黑箱模型,常常因编造事实知识、缺乏可解释性备受批评,面临着人工智能幻觉、专业领域实用性差等重大挑战,严重影响了大语言模型在军事决策、医疗诊断和法律判决等高风险、高技能领域的应用,多模态知识图谱存储了类型多样且准确的事实知识,可以通过提供外部知识完成推理和解释[134]。不同模型和领域有各自的需求和特性,面临着不同的挑战,因此利用大模型与多模态知识图谱的协同作用来实现多模态知识图谱在高风险、高技能领域的应用是非常有必要的。未来,通过多模态知识图谱和大模型的交互融合来实现更智能、更强大的智能系统应用,打通面向通用人工智能的新途径,实现人工智能在各领域落地的新模式,因此实现多模态知识图谱与大语言模型的交互融合,将会是多模态知识图谱发展的一个新方向。

相对于社会其他领域,军事领域数据还面临着标注数据极度缺乏、数据具有不完整性、存在大量数据噪声与数据欺诈、因保密性而存在的交互性差和更新迭代慢等挑战,并且军事领域需要一个面向全军兵种的军事数据综合体来整合、应用海量异构数据。下面提出一些未来多模态知识图谱应用于军事领域可关注的方向。

(1)军事数据存在数据量少和标注数据稀缺问题,零次学习、小样本学习、迁移学习以及元学习等模型不依赖大量数据,通过在类别相关的非军事训练集的训练,应用到军事领域,以减少对数据的依赖。

(2)大量军事数据因战场环境的复杂性存在遮挡、缺失等不完整性的挑战,生成对抗网络(generative adversarial network,GAN)、深度卷积生成对抗网络(deep convolutional generative adversarial network,DCGAN)、掩码自编码器(MAE)等方法可以根据缺失数据补全为完整数据,针对情报侦察中获得的不完整战场数据进行信息补全获得完整的战场数据,为多模态知识图谱构建提供更准确、可靠的数据源。

(3)随着认知战与舆论战的出现,军事情报与社交媒体中出现了越来越多的虚假数据,通过数据欺诈的手段干扰敌方的认知决策与敌方群众认知。未来可根据分析以往欺诈数据的发出者、数据形式、数据内容、媒体特征、传播方式和目标受众等诸多要素,开发欺骗检测框架,并部署在军事多模态知识图谱构建的多个步骤,层层分析,以实现欺诈数据监测与销毁。

(4)军事数据因保密性只能在内网流转,缺乏与外界的交互,造成了数据更新迭代慢,并且编制体制呈现树状结构,各军兵种专业繁杂,数据互通性差。未来可以通过多模态知识图谱与大模型的交互融合,结合军事领域特点,构建仅面向军事领域的多模态知识图谱,部署在统一的军队综合网上形成面向全军兵种的军事数据综合体,通过数据的实时上传,对不同等级的使用者设置不同的使用权限,对不同密级的军事数据设置不同的调阅权限,实现远程数据调阅、实时数据分析、战场态势感知、远程作战支持、智能辅助决策等下游应用,提升数据更新速度,打破各军兵种不同业务领域间的数据壁垒,提高整体作战效率。

5 结束语

在大模型被广泛使用以及数据资源呈现爆炸式增长的背景下,多模态知识图谱成为一个新的研究热点。本文对多模态知识图谱的构建技术及在军事领域的应用展开综述,汇总了现有多模态知识图谱以及数据资源,分析了多模态信息抽取、多模态实体链接以及多模态表示学习三种多模态知识图谱构建的关键技术,以及该技术领域的研究现状和代表性方法,创新性地总结了大模型技术在多模态知识图谱构建过程中的运用。根据应用方向,探究了多模态知识图谱在军事领域中的重要应用。结合大模型的热点话题和军事需求,对多模态知识图谱在构建技术与军事应用中的发展方向提出了展望,希望能够为多模态知识图谱的构建及在军事领域的进一步发展提供有效的理论支撑与实际参考。

参考文献:

[1] 谢作如. 用PyWebIO“交互”呈现人工智能学习成果[J]. 中国信息技术教育, 2021(15): 82-84.
XIE Z R. Using PyWebIO “interaction” to present artificial intelligence learning results[J]. Chinese Information Technology Education, 2021(15): 82-84.

[2] SINGHAL A. Introducing the knowledge graph: things, not strings[EB/OL]. [2021-11-19]. https://blog.google/products/search/introducing-knowledge-graph-things-not/.

[3] 李涓子, 侯磊. 知识图谱研究综述[J]. 山西大学学报 (自然科学版), 2017, 40(3): 454-459.
LI J Z, HOU L. Reviews on knowledge graph research[J]. Journal of Shanxi University (Natural Science Edition), 2017, 40(3): 454-459.

[4] O'CALLAGHAN J. How OpenAI's text-to-video tool Sora could change science-and society[J]. Nature, 2024, 627(8004): 475-476.

[5] XU G, CHEN H, LI F L, et al. AliME MKG: a multi-modal knowledge graph for live-streaming e-commerce[C]//Proceedings of the 30th ACM International Conference on Information and Knowledge Management, 2021: 4808-4812.

[6] ZHA Z, WANG J, LI Z, et al. M2ConceptBase: a fine-grained aligned multi-modal conceptual knowledge base[J]. arXiv:2312.10417, 2023.

[7] ZHU X, LI Z, WANG X, et al. Multi-modal knowledge graph construction and application: a survey[J]. IEEE Transactions

on Knowledge and Data Engineering, 2024, 36(2): 715-735.
[8] PENG J, HU X, HUANG W, et al. What is a multi-modal knowledge graph: a survey[J]. Big Data Research, 2023, 32: 100380.
[9] CHEN Y, GE X, YANG S, et al. A survey on multimodal knowledge graphs: construction, completion and applications [J]. Mathematics, 2023, 11(8): 1815.
[10] 陈烨, 周刚, 卢记仓. 多模态知识图谱构建与应用研究综述[J]. 计算机应用研究, 2021, 38(12): 3535-3543.
CHEN Y, ZHOU G, LU J C. Survey on construction and application research for multi-modal knowledge graphs[J]. Application Research of Computers, 2021, 38(12): 3535-3543.
[11] 陈佳云, 徐向英, 章永龙, 等. 多模态知识图谱在农业中的研究进展[J]. 农业大数据学报, 2022, 4(3): 126-134.
CHEN J Y, XU X Y, ZHANG Y L, et al. Research progress of multimodal knowledge graph in agriculture[J]. Journal of Agricultural Big Data, 2022, 4(3): 126-134.
[12] WANG M, QI G, WANG H F, et al. RichPedia: a comprehensive multi-modal knowledge graph[C]//Proceedings of the 9th Joint International Conference on Semantic Technology, Hangzhou, Nov 25-27, 2019: 130-145.
[13] DENG J, DONG W, SOCHER R, et al. ImageNet: a large-scale hierarchical image database[C]//Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009: 248-255.
[14] LEHMANN J, ISELE R, JAKOB M, et al. DBPedia—a large-scale, multilingual knowledge base extracted from WikiPedia [J]. Semantic Web, 2015, 6(2): 167-195.
[15] WU Y, WU X, LI J, et al. MMPedia: a large-scale multi-modal knowledge graph[C]//Proceedings of the 2023 International Semantic Web Conference. Cham: Springer, 2023: 18-37.
[16] ZHANG J, WANG J, WANG X, et al. AspectMMKG: a multi-modal knowledge graph with aspect-aware entities[C]//Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, 2023: 3361-3370.
[17] XU B, XU Y, LIANG J, et al. CN-DBpedia: a never-ending Chinese knowledge extraction system[C]//Proceedings of the 2017 International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems. Cham: Springer, 2017: 428-438.
[18] FERRADA S, BUSTOS B, HOGAN A. IMGPedia: a linked dataset with content-based analysis of WikiMedia images[C]//Proceedings of the 16th International Semantic Web Conference, Vienna, Oct 21-25, 2017. Cham: Springer, 2017: 84-93.
[19] OÑORO-RUBIO D, NIEPERT M, GARCÍA-DURÁN A, et al. Answering visual-relational queries in web-extracted knowledge graphs[J]. arXiv:1709.02314, 2017.
[20] ZHANG N, LI L, CHEN X, et al. Multimodal analogical reasoning over knowledge graphs[J]. arXiv:2210.00312, 2022.
[21] 李华昱, 付亚凤, 闫阳, 等. 基于LEBERT的多模态领域知识图谱构建[J]. 计算机系统应用, 2022, 31(11): 79-90.
LI H Y, FU Y F, YAN Y et al. Construction of multi-modal domain knowledge map based on LEBERT[J]. Computer Systems & Applications, 2022, 31(11): 79-90.
[22] ZHENG S, WANG W, QU J, et al. MMKGR: multi-hop multi-modal knowledge graph reasoning[C]//Proceedings of the 2023 IEEE 39th International Conference on Data Engineering, 2023: 96-109.
[23] XU G, MENG Y, QIU X, et al. Sentiment analysis of comment texts based on BiLSTM[J]. IEEE Access, 2019, 7: 51522-51532.
[24] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding [J]. arXiv:1810.04805, 2018.
[25] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
[26] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[C]//Advances in Neural Information Processing Systems 25, 2012.
[27] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481-2495.
[28] HE K, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]//Proceedings of the 2017 IEEE International Conference on Computer Vision, 2017: 2961-2969.
[29] CHEN L C, PAPANDREOU G, KOKKINOS I, ET al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(4): 834-848.
[30] SINGH A, LIU H, PLUMBLEY M D. E-PANNs: sound recognition using efficient pre-trained audio neural networks [J]. arXiv:2305.18665, 2023.
[31] CHAN W, JAITLY N, LE Q, et al. Listen, attend and spell: a neural network for large vocabulary conversational speech recognition[C]//Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing, 2016: 4960-4964.
[32] ZHAO Q, GAO T, GUO N. TSVFN: two-stage visual fusion network for multimodal relation extraction[J]. Information Processing & Management, 2023, 60(3): 103264.
[33] WU J, CUI Z, SHENG V S, et al. A comparative study of SIFT and its variants[J]. Measurement Science Review, 2013, 13(3): 122-131.
[34] KADOTA R, SUGANO H, HIROMOTO M, et al. Hardware architecture for HOG feature extraction[C]//Proceedings of the 2009 5th International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2009: 1330-1333.
[35] OYALLON E, RABIN J. An analysis of the SURF method [J]. Image Processing on Line, 2015, 5: 176-218.

[36] RUBLEE E, RABAUD V, KONOLIGE K, et al. ORB: an efficient alternative to SIFT or SURF[C]//Proceedings of the 2011 International Conference on Computer Vision, 2011: 2564-2571.

[37] WANG X, ZOU J, SHI D. An improved ORB image feature matching algorithm based on SURF[C]//Proceedings of the 2018 3rd International Conference on Robotics and Automation Engineering, 2018: 218-222.

[38] ZHANG H Q, ZHANG J L, DAI R Y. A fast image matching research based on MIC-SURF algorithm[C]//Proceedings of the 27th Chinese Control and Decision Conference, 2015: 542-547.

[39] ZHANG D, WEI S, LI S, et al. Multi-modal graph fusion for named entity recognition with targeted visual guidance[C]// Proceedings of the 35th AAAI Conference on Artificial Intelligence, 2021: 14347-14355.

[40] ZHANG Q, FU J, LIU X, et al. Adaptive co-attention network for named entity recognition in tweets[C]//Proceedings of the 32nd AAAI Conference on Artificial Intelligence, 2018.

[41] LU D, NEVES L, CARVALHO V, et al. Visual attention model for name tagging in multimodal social media[C]// Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2018: 1990-1999.

[42] RIAZ F, ANWAR M W, MUQADES H. Maximum entropy based URDU named entity recognition[C]//Proceedings of the 2020 International Conference on Engineering and Emerging Technologies, 2020: 1-5.

[43] PATIL N V, PATIL A S, PAWAR B V. HMM based named entity recognition for inflectional language[C]//Proceedings of the 2017 International Conference on Computer, Communications and Electronics, 2017: 565-572.

[44] TANG B, FENG Y, WANG X, et al. A comparison of conditional random fields and structured support vector machines for chemical entity recognition in biomedical literature[J]. Journal of Cheminformatics, 2015, 7(1): 1-6.

[45] LIU M, TU Z, ZHANG T, et al. LTP: a new active learning strategy for CRF-based named entity recognition[J]. Neural Processing Letters, 2022, 54(3): 2433-2454.

[46] GU J, WANG Z, KUEN J, et al. Recent advances in convolutional neural networks[J]. Pattern Recognition, 2018, 77: 354-377.

[47] XU X, XU J, RUAN G, et al. A pipeline-based multimodal military event argument extraction framework[C]//Proceedings of the 2022 China Conference on Knowledge Graph and Semantic Computing. Singapore: Springer, 2022: 21-29.

[48] WANG X, YE J, LI Z, et al. CAT-MNER: multimodal named entity recognition with knowledge-refined cross-modal attention[C]//Proceedings of the 2022 IEEE International Conference on Multimedia and Expo, 2022: 1-6.

[49] SUN L, WANG J, ZHANG K, et al. RpBERT: a text-image relation propagation-based BERT model for multimodal NER [C]//Proceedings of the 35th AAAI Conference on Artificial Intelligence, 2021: 13860-13868.

[50] ZHANG X, YUAN J, LI L, et al. Reducing the bias of visual objects in multimodal named entity recognition[C]//Proceedings of the 16th ACM International Conference on Web Search and Data Mining, 2023: 958-966.

[51] WANG P, WU J, CHEN X. Multimodal entity linking with gated hierarchical fusion and contrastive training[C]//Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2022: 938-948.

[52] WANG J, YANG Y, MAO J, et al. CNN-RNN: a unified framework for multi-label image classification[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016: 2285-2294.

[53] YOO J, KIM T, LEE S, et al. Enriched CNN-transformer feature aggregation networks for super-resolution[C]//Proceedings of the 2023 IEEE/CVF Winter Conference on Applications of Computer Vision, 2023: 4956-4965.

[54] HAORAN X, ZIYI W. Condition evaluation and fault diagnosis of power transformer based on GAN-CNN[J]. Journal of Electrotechnology, Electrical Engineering and Management, 2023, 6(3): 8-16.

[55] DHIRAVIDACHELVI E, PRABAVATHI R. Artificial humming bird optimization-based hybrid CNN-RNN for accurate exudate classification from fundus images[J]. Journal of Digital Imaging, 2023, 36(1): 59.

[56] CHEN B, ZOU X, ZHANG Y, et al. LEFormer: a hybrid CNN-Transformer architecture for accurate lake extraction from remote sensing imagery[C]//Proceedings of the 2024 IEEE International Conference on Acoustics, Speech and Signal Processing, 2024: 5710-5714.

[57] PARSEH M J, RAHMANIMANESH M, KESHAVARZI P, et al. Scene representation using a new two-branch neural network model[J]. The Visual Computer, 2024, 40: 6219-6244.

[58] SUBRAHMANYESWARA R B. Accurate Leukocoria predictor based on deep VGG-net CNN technique[J]. IET Image Processing, 2020, 14(10): 2241-2248.

[59] ANAND R, SHANTHI T, NITHISH M S, et al. Face recognition and classification using GoogleNET architecture[C]// Soft Computing for Problem Solving, Vellore, Dec 17-19, 2018. Singapore: Springer, 2020: 261-269.

[60] ZHANG K, GUO Y, WANG X, et al. Multiple feature reweight DenseNet for image classification[J]. IEEE Access, 2019, 7: 9872-9880.

[61] HOWARD A G, ZHU M, CHEN B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[J]. arXiv:1704.04861, 2017.

[62] TAN M, LE Q. EfficientNet: rethinking model scaling for convolutional neural networks[C]//Proceedings of the 2019 International Conference on Machine Learning, 2019: 6105-

6114.
[63] WU S, FEI H, CAO Y, et al. Information screening whilst exploiting! Multimodal relation extraction with feature denoising and multimodal topic modeling[J]. arXiv:2305.11719, 2023.
[64] LU Y, YANG R, JIANG X, et al. MRE: a military relation extraction model based on BiGRU and multi-head attention [J]. Symmetry, 2021, 13(9): 1742.
[65] YU J, LI Z, WANG J, et al. Grounded multimodal named entity recognition on social media[C]//Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2023: 9141-9154.
[66] LEE C E, BAEK J, SON J, et al. Deep AI military staff: cooperative battlefield situation awareness for commander's decision making[J]. The Journal of Supercomputing, 2023, 79(6): 6040-6069.
[67] 余晓晗, 毛绍臣, 綦秀利, 等. 基于改进MAE的装甲车辆目标前景遮挡部分补全方法[J]. 火力与指挥控制, 2023, 48(11): 72-80.
YU X H, MAO S C, QI X L, et al. Foreground occlusion complementation method for armored vehicle targets based on improved masked auto encoder (MAE)[J]. Fire Control & Command Control, 2023, 48(11): 72-80.
[68] PARCHAMI M, ZHU W P, CHAMPAGNE B, et al. Recent developments in speech enhancement in the short-time Fourier transform domain[J]. IEEE Circuits and Systems Magazine, 2016, 16(3): 45-77.
[69] SRIVASTAVA R K, PANDEY D. Speech recognition using HMM and soft computing[J]. Materials Today: Proceedings, 2022, 51: 1878-1883.
[70] GUPTA H, GUPTA D. LPC and LPCC method of feature extraction in speech recognition system[C]//Proceedings of the 2016 6th International Conference on Cloud System and Big Data Engineering (Confluence), 2016: 498-502.
[71] HIDAYAT R. Frequency domain analysis of MFCC feature extraction in children's speech recognition system[J]. Jurnal Infotel, 2022, 14(1): 30-36.
[72] NUGROHO H, FUADIYAH R N N. Development of speech emotion recognition system based on discrete wavelet transform (DWT) and voice segmentation[J]. International Journal on Electrical Engineering and Informatics, 2022, 14(3): 593-607.
[73] LABIED M, BELANGOUR A. Automatic speech recognition features extraction techniques: a multi-criteria comparison [J]. International Journal of Advanced Computer Science and Applications, 2021, 12(8).
[74] PRABHAVALKAR R, HORI T, SAINATH T N, et al. End-to-end speech recognition: a survey[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2024, 32: 325-351.
[75] WU T, WANG G, ZHAO J, et al. Towards relation extraction from speech[J]. arXiv:2210.08759, 2022.
[76] GHANNAY S, CAUBRIERE A, ESTEVE Y, et al. End-to-end named entity extraction from speech[J]. arXiv:1805.12045, 2018.
[77] CHEN B, XU G, WANG X, et al. Aishell-NER: named entity recognition from Chinese speech[C]//Proceedings of the 2022 IEEE International Conference on Acoustics, Speech and Signal Processing, 2022: 8352-8356.
[78] LI J, LI H, SUN D, et al. LLMs as bridges: reformulating grounded multimodal named entity recognition[J]. arXiv: 2402.09989, 2024.
[79] CHEN F, FENG Y. Chain-of-thought prompt distillation for multimodal named entity and multimodal relation extraction [J]. arXiv:2306.14122, 2023.
[80] CHEN G, SHEN L, SHAO R, et al. LION: empowering multimodal large language model with dual-level visual knowledge[C]//Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 26540-26550.
[81] WANG X, ZHOU W, ZU C, et al. InstructUIE: multi-task instruction tuning for unified information extraction[J]. arXiv: 2304.08085, 2023.
[82] HE W, MA H, LI S, et al. Using augmented small multimodal models to guide large language models for multimodal relation extraction[J]. Applied Sciences, 2023, 13(22): 12208.
[83] 王震宇, 朱学芳, 杨睿. 基于多模态大语言模型的关系抽取研究[J/OL]. 数据分析与知识发现 [2024-05-11]. http://kns.cnki.net/kcms/detail/10.1478.G2.20240117.1110.024.html.
WANG Z Y, ZHU X F, YANG R. Research on relation extraction based on multimodal large language model[J/OL]. Data Analysis and Knowledge Discovery [2024-05-11]. http://kns.cnki.net/kcms/detail/10.1478.G2.20240117.1110.024.html.
[84] CUI H, FANG X, ZHANG Z, et al. Open visual knowledge extraction via relation-oriented multimodality model prompting [C]//Advances in Neural Information Processing Systems 36, 2024.
[85] CHENG Z D, JU C, CHEN X, et al. Image to multi-modal retrieval for industrial scenarios[J]. arXiv:2305.03972, 2023.
[86] GULZAR Y, ALWAN A A, ABDULLAH R M, et al. OCA: ordered clustering-based algorithm for e-commerce recommendation system[J]. Sustainability, 2023, 15(4): 2947.
[87] ZHOU C, MA J, ZHANG J, et al. Contrastive learning for debiased candidate generation in large-scale recommender systems[C]//Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2021: 3985-3995.
[88] SUI X, ZHANG Y, SONG K, et al. Improving zero-shot entity linking candidate generation with ultra-fine entity type information[C]//Proceedings of the 29th International Conference on Computational Linguistics, 2022: 2429-2437.
[89] LUO P, XU T, WU S, et al. Multi-grained multimodal interaction network for entity linking[C]//Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2023: 1583-1594.

[90] YANG C, HE B, WU Y, et al. MMEL: a joint learning framework for multi-mention entity linking[C]//Proceedings of Machine Learning Research 216: Uncertainty in Artificial Intelligence, Pittsburgh, 2023: 2411-2421.

[91] GAN J, LUO J, WANG H, et al. Multimodal entity linking: a new dataset and a baseline[C]//Proceedings of the 29th ACM International Conference on Multimedia, 2021: 993-1001.

[92] HUANG H, HECK L, JI H. Leveraging deep neural networks and knowledge graphs for entity disambiguation[J]. arXiv: 1504.07678, 2015.

[93] MOON S, NEVES L, CARVALHO V. Multimodal named entity disambiguation for noisy social media posts[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2018: 2000-2008.

[94] ZHOU P, YING K, WANG Z, et al. Self-supervised enhancement for named entity disambiguation via multimodal graph convolution[J]. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(1): 231-245.

[95] ZHANG D, HUANG L. Multimodal knowledge learning for named entity disambiguation[C]//Findings of the Association for Computational Linguistics: EMNLP 2022, Abu Dhabi, 2022: 3160-3169.

[96] YAO B M, CHEN Y, WANG Q, et al. AMELI: enhancing multimodal entity linking with fine-grained attributes[J]. arXiv:2305.14725, 2023.

[97] ZHAO H, WANG X, CHEN S, et al. OVEL: large language model as memory manager for online video entity linking [J]. arXiv:2403.01411, 2024.

[98] YANG L, CHEN H, WANG X, et al. Two heads are better than one: integrating knowledge from knowledge graphs and large language models for entity alignment[J]. arXiv:2401.16960, 2024.

[99] SHI S, XU Z, HU B, et al. Generative multimodal entity linking[J]. arXiv:2306.12725, 2023.

[100] ROSSETTO L, KYRIAKOU A, LANGE S, et al. LifeGraph 4-lifelog retrieval using multimodal knowledge graphs and vision-language models[C]//Proceedings of the 7th Annual ACM Workshop on the Lifelog Search Challenge, 2024: 88-92.

[101] LONG X, ZENG J, MENG F, et al. Generative multi-modal knowledge retrieval with large language models[C]//Proceedings of the 38th AAAI Conference on Artificial Intelligence, 2024: 18733-18741.

[102] XIAO Z, GONG M, CASCANTE-BONILLA P, et al. Grounding language models for visual entity recognition [J]. arXiv:2402.18695, 2024.

[103] SATO M, KOBAYASHI T, SOROIDA Y, et al. Development of novel deep multimodal representation learning-based model for the differentiation of liver tumors on B-mode ultrasound images[J]. Journal of Gastroenterology and Hepatology, 2022, 37(4): 678-684.

[104] HUANG Z, XU X, NI J, et al. Multimodal representation learning for recommendation in Internet of things[J]. IEEE Internet of Things Journal, 2019, 6(6): 10675-10685.

[105] YU L, CHEN J, SINHA A, et al. CommerceMM: large-scale commerce multimodal representation learning with omni retrieval[C]//Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2022: 4433-4442.

[106] HU P, ZHEN L, PENG D, et al. Scalable deep multimodal learning for cross-modal retrieval[C]//Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2019: 635-644.

[107] QI J, PENG Y. Cross-modal bidirectional translation via reinforcement learning[C]//Proceedings of the 27th International Joint Conference on Artificial Intelligence, 2018: 2630-2636.

[108] ZHOU F, CHEN H. Cross-modal translation and alignment for survival analysis[C]//Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision, 2023: 21485-21494.

[109] YE R, WANG M, LI L. End-to-end speech translation via cross-modal progressive training[J]. arXiv:2104.10380, 2021.

[110] SHAO H, QIAN S, XIAO H, et al. Visual CoT: unleashing chain-of-thought reasoning in multi-modal language models [J]. arXiv:2403.16999, 2024.

[111] LI X, LIAN D, LU Z, et al. GraphAdapter: tuning vision-language models with dual knowledge graph[C]//Advances in Neural Information Processing Systems 36, 2024.

[112] MONDAL D, MODI S, PANDA S, et al. KAM-CoT: knowledge augmented multimodal chain-of-thoughts reasoning [C]//Proceedings of the 38th AAAI Conference on Artificial Intelligence, 2024: 18798-18806.

[113] CHAWLA R, DATTA A, VERMA T, et al. Veagle: advancements in multimodal representation learning[J]. arXiv:2403.08773, 2024.

[114] LI W, FAN H, WONG Y, et al. Improving context understanding in multimodal large language models via multimodal composition learning[C]//Proceedings of the 41st International Conference on Machine Learning, 2024.

[115] TAI Y, FAN W, ZHANG Z, et al. Link-context learning for multimodal LLMs[C]//Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 27176-27185.

[116] WEI J, WANG X, SCHUURMANS D, et al. Chain-of-thought prompting elicits reasoning in large language models [C]//Advances in Neural Information Processing Systems 35, 2022: 24824-24837.

[117] 蔡群, 付美玲, 黎文, 等. 电子目标知识图谱构建及运用[C]//中国指挥与控制学会. 第十一届中国指挥控制大会论文集, 2023: 6.
CAI Q, FU M L, LI W, et al. Construction and application of electronic target knowledge graph[C]//Chinese Institute of Command and Control. Proceedings of the 11th China

Conference on Command and Control, 2023: 6.

[118] 顾丹阳, 李明倩, 权冀川, 等. 基于本体的主战武器装备知识图谱构建[J]. 指挥控制与仿真, 2021, 43(6): 14-20.
GU D Y, LI M Q, QUAN J C, et al. Ontology based knowledge graph construction for combat weapon equipment[J]. Command Control & Simulation, 2021, 43(6): 14-20.

[119] 黄伟春, 肖刚, 杨健, 等. 基于本体的军事术语知识图谱构建方法[J]. 指挥控制与仿真, 2023, 45(5): 10-17.
HUANG W C, XIAO G, YANG J, et al. Ontology-based military terminology knowledge graph construction method [J]. Command Control & Simulation, 2023, 45(5): 10-17.

[120] 邢萌, 杨朝红, 毕建权. 军事领域知识图谱的构建及应用[J]. 指挥控制与仿真, 2020, 42(4): 1-7.
XING M, YANG C H, BI J Q. Construction and application of domain-specific knowledge graph in military field [J]. Command Control & Simulation, 2019, 42(4): 1-7.

[121] 袁清波, 杜晓明, 姚奕, 等. 融合汉字多特征的指挥控制保障领域命名实体识别[J]. 火力与指挥控制, 2022, 47(9): 48-53.
YUAN Q B, DU X M, YAO Y, et al. Named entity recognition in command and control support domain based on multi feature of Chinese characters[J]. Fire Control & Command Control, 2022, 47(9): 48-53.

[122] WANG Y, WANG T, WANG J, et al. Military chain: construction of domain knowledge graph of kill chain based on natural language model[J]. Mobile Information Systems, 2022.

[123] 傅浩, 刘姗姗, 李佳蔚, 等. 军事事件主题检测与抽取方法[J]. 指挥信息系统与技术, 2024, 15(1): 76-81.
FU H, LIU S S, LI J W, et al. Military event theme detection and extraction method[J]. Command Information System and Technology, 2024, 15(1): 76-81.

[124] 成浩, 梁平, 刘超鑫, 等. 情报信息驱动下的军事目标知识深度认知研究[J]. 网络安全与数据治理, 2023, 42(S2): 139-143.
CHENG H, LIANG P, LIU C X, et al. Research on the construction of military target knowledge graph based on intelligence information driven[J]. Cyber Security and Data Governance, 2023, 42(S2): 139-143.

[125] 贺玲, 贺照辉. 大数据技术在战场态势感知中的应用[J]. 科技与创新, 2023(7): 178-181.
HE L, HE Z H. Application of big data technology in battlefield situational awareness[J]. Science and Technology & Innovation, 2023(7): 178-181.

[126] 王昊奋, 易侃, 吴蔚, 等. 多模态态势感知的知识表示、表示学习和知识推理[J]. 指挥信息系统与技术, 2022, 13(3): 1-11.
WANG H F, YI K, WU W, et al. Knowledge representation, representation learning and knowledge reasoning for multi-modal situational awareness[J]. Command Information System and Technology, 2022, 13(3): 1-11.

[127] 黄梓航, 蒋秉川, 刘靖旭. 战场环境知识图谱智能服务系统设计和关键技术研究[J]. 军事运筹与系统工程, 2021, 35(4): 73-80.
HUANG Z H, JIANG B C, LIU J X. Battlefield environment knowledge map intelligent service system design and key technology research[J]. Military Operations Research and Systems Engineering, 2021, 35(4): 73-80.

[128] 时空知识图谱应用初探[EB/OL]. (2021-12-21) [2024-03-10]. https://blog.csdn.net/bmy0000/article/details/122055190.
Application of spatiotemporal knowledge graph[EB/OL]. (2021-12-21) [2024-03-10]. https://blog.csdn.net/bmy0000/article/details/122055190.

[129] 彭京徽, 汪振, 李越, 等. 装备领域多模态知识图谱技术研究[J]. 兵器装备工程学报, 2022, 43(11): 136-140.
PENG J H, WANG Z, LI Y, et al. Research on multimodal knowledge graph technology in equipment field[J]. Journal of Ordnance Equipment Engineering, 2022, 43(11): 136-140.

[130] 胡卫, 赵文龙, 李石磊, 等. 军事装备管理数据知识图谱构建及应用[J]. 火力与指挥控制, 2022, 47(10): 125-131.
HU W, ZHAO W L, LI S L, et al. Research on the construction and application of knowledge graph of military equipment management data[J]. Fire Control & Command Control, 2022, 47(10): 125-131.

[131] 王宏宇, 许潇, 周育伟, 等. 基于军事领域知识图谱的智能问答系统设计与实现[J]. 装甲兵学报, 2022, 1(2): 87-94.
WANG H Y, XU X, ZHOU Y W, et al. Design and implementation of intelligent question-and-answer system based on knowledge graph in military field[J]. Journal of Armored Forces, 2022, 1(2): 87-94.

[132] 宗滕, 吴松涛, 周春华. 基于多模态数据分析的典型智能化军事应用[J]. 信息安全与通信保密, 2022(2): 9-16.
ZONG T, WU S T, ZHOU C H. Typical intelligent military application based on multimodal data analysis[J]. Information Security and Communications Privacy, 2022(2): 9-16.

[133] 李卫星, 王峰, 李智国, 等. 面向多源数据的军事信息系统设计[J]. 中国电子科学研究院学报, 2020, 15(3): 237-243.
LI W X, WANG F, LI Z G, et al. Design of military information system based on multi- source data[J]. Journal of China Academy of Electronics and Information Technology, 2020, 15(3): 237-243.

[134] PAN S, LUO L, WANG Y, et al. Unifying large language models and knowledge graphs: a roadmap[J]. arXiv:2306.08302, 2023.