

## 事件提取方法在军事领域的应用趋势

吴 蕾<sup>1,2</sup>, 邓牲岫<sup>1,3</sup>, 柳少军<sup>1</sup>, 李志强<sup>1</sup>

(1.国防大学联合作战学院, 北京 100091; 2.陆军航空兵研究所, 北京 101121; 3.陆军工程大学, 江苏 南京 210001)

**摘 要:**事件提取可以帮助用户从海量、无序的非结构化信息中快速、准确地获取感兴趣的事件,在自然语言处理领域有广泛应用。在梳理事件的概念、知识表示以及事件提取发展历程的基础上,对元事件和主题事件的提取方法分别进行了归纳和分析,并总结了事件提取方法在军事上的研究现状,探讨了其在军事领域未来的应用趋势。

**关键词:**事件提取; 元事件; 主题事件; 机器学习; 深度学习

**中图分类号:**E911

**文献标志码:**A

**DOI:**10.3969/j.issn.1673-3819.2021.06.022

### Event Extraction Methods and Development Trend in Military Field

WU Lei<sup>1,2</sup>, DENG Shen-shen<sup>1,3</sup>, LIU Shao-jun<sup>1</sup>, LI Zhi-qiang<sup>1</sup>

(1. Joint Operations College, National Defence University, Beijing 100091; 2. Army Aviation Research Institute, Beijing 101121; 3. Army Engineering University, Nanjing 210001, China)

**Abstract:** Event extraction method can help users quickly and accurately identify the interesting events from the massive, disordered and unstructured information, which is widely used in the field of natural language processing. On the basis of sorting out the concept, the knowledge representation of event and the development of event extraction, the extraction methods of meta-event and topic event are summarized and analyzed respectively. The research status of event extraction methods in military field is summarized, and the application trend of event extraction methods in military field is discussed.

**Key words:** event extraction; meta-event; topic event; machine learning; deep learning

大型计算机模拟对抗演习是和平时训练指挥员作战指挥能力的主要手段,通常涵盖陆、海、空、天、电、网多维战场空间,具有模拟层次高、仿真规模大、覆盖范围广、演习要素全、描述实体多、行动交互复杂等特点。

对抗演习过程中,人与计算机模拟系统交互作用产生海量的仿真信息,这些信息通常包括作战计划、行动指令、导调文书等内容,涵盖结构化、半结构化以及非结构化多种形式,其多样性和复杂性给指挥员带来了巨大的认知压力。因此,迫切需要一种方法,帮助指挥员从这些海量信息中快速、准确地筛选出较为关键的事件,为指挥决策或者检验评估提供支撑。

目前,国内外针对事件提取研究主要集中在元事件提取方面,主题事件提取也逐渐引起关注,但事件提取方法在军事领域应用尚不成熟。本文对事件的概念、事件提取的发展历程进行了分析,归纳梳理了当前常用的事件提取方法以及各自的特点和局限,并结合事件提取在军事领域研究现状和技术发展,指出了其在军事领域未来的应用趋势,为进一步研究相关领域事件提取提供参考。

## 1 事件提取相关概念及发展历程

### 1.1 事件和事件提取

事件(Event)的概念来源于认知科学,相关研究者认为人类的记忆由事件以及事件之间的关系构成。随后,事件的概念逐渐发展到其他领域。

在信息检索与信息提取领域,事件一般以句子为单位,指在特定时间段和特定区域内发生的事情,涉及角色的参与,并且由行动组成。Allan 等认为事件是“细化了的用于检索的主题”<sup>[1]</sup>,Yang 等将事件定义为“在一个特定时间、特定地点发生的事情”<sup>[2]</sup>。

在自动文本摘要领域,事件是比参与者、时间和地点等概念具有更大粒度的语义单元,具有动态性和完整意义。杨竣辉<sup>[3]</sup>将事件作为最基本的语义单元,通过研究事件及事件间的关系来表示文本的语义。王伟玉等<sup>[4]</sup>提出了一种事件粒度的话题表示方法,认为通过融合事件描述的共性信息,可以生成事件粒度的话题的简短表示。

事件提取方法主要研究如何从描述事件信息的数据或语句中提取事件信息并以结构化的形式呈现出来,包括事件时间、事件地点、参与者以及动作或状态的变化等事件要素。

### 1.2 事件提取发展历程

事件提取方法取得巨大进步,与测评会议 MUC、TDT 及 ACE 的推动密不可分,这些测评会议虽然研究

收稿日期:2021-05-05

修回日期:2021-06-09

**作者简介:**吴蕾(1983—),女,江苏泰兴人,博士研究生,工程师,研究方向为运筹分析与军事智能决策。  
邓牲岫(1982—),女,博士研究生。

的侧重点各有不同,但是它们的召开在很大程度上促进了事件提取方法的快速发展。

事件提取研究来源于20世纪80年代美国国防部高级研究计划局(Defense Advanced Research Projects Agency, DARPA)主办的消息理解会议(Message Understanding Conference, MUC)。随着信息化战争的到来,军事数据数量巨大并且飞速增加,从纷繁复杂的数据中提取关键信息就显得极为重要。因此,会议最早的语料来源是美军的作战文书,任务是从这些作战文书中抽取相关事件,填入预先设置的模板。这一系列会议的召开标志着信息提取开始成为自然语言处理领域的一个重要分支。

另一个评测会议——话题识别与跟踪(Topic Detection and Tracking, TDT)会议也是由DARPA主办的,它的主要任务是对面向新闻信息事件识别和提取的技术进行研究和评测,目的是通过对文本的划分、对新闻信息流的监控以及对同一话题下的分散报道的有效组织,发现特定领域新事件的报道。会议初期,学者们指出一个话题就是一个特定事件,随着会议的召开,话题逐渐发展为相互之间有关系的多个事件的组合。

由美国国家标准与技术研究所(National Institute of Standards and Technology, NIST)组织的自动内容抽取(Automatic Content Extraction, ACE)国际测评会议进一步推动了事件提取研究的发展,这是事件提取领域非常重要的系列会议,主要研究如何从新闻语料库中自动抽取实体、关系、事件等内容。与MUC会议相比,ACE会议不针对具体的领域或场景,也不预先设置模板,更强调对文本中事件要素的识别与描述。

## 2 事件提取的分类

元事件表示一个动作的发生或状态的变化,它是主题事件的基本组成单位。目前国内外学者对于元事件提取的研究已经比较成熟,对于主题事件提取的研究也越来越重视。

### 2.1 元事件提取

随着事件提取技术的发展,元事件提取先后出现三种主要方法。

最早出现的是基于模式匹配的事件提取方法,其在提取事件时用模式进行约束,从而找出符合约束条件的事件,具体提取流程如图1所示。国外很早就开展了这方面的研究工作,并陆续开发了PALK、TIMES、AutoSlog-TS等基于模式匹配的事件提取系统。国内事件提取研究开展较晚,研究者们陆续定义了一些事件的模式并提出了相关模式学习方法,这些方法主要是利用与领域无关的知识库进行模式学习,进而

实现事件提取。

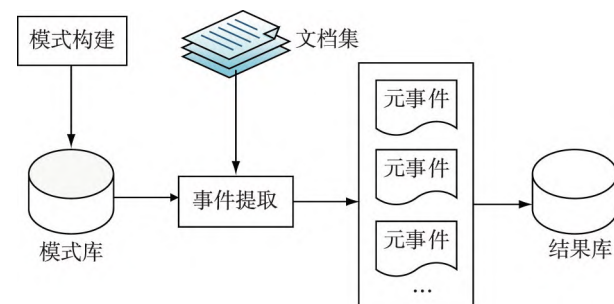


图1 基于模式匹配的元事件提取流程

将传统机器学习方法应用到事件提取中主要是通过特征选择,训练出分类效果较好的分类器,进而实现事件的提取。用于事件提取的传统机器学习方法主要有支持向量机(Support Vector Machine, SVM)模型、最大熵(Maximum Entropy, ME)模型、隐马尔可夫模型(Hidden Markov Model, HMM)、条件随机场(Conditional Random Fields, CRF)模型等方法,它们各有其局限性,比如,SVM模型难以在大规模训练样本上进行,HMM需要严格的独立性假设作为前提,ME模型迭代过程计算量非常大,而CRF模型较复杂且训练代价较高。

随着人工智能技术的飞速发展,以循环神经网络(Recurrent Neural Network, RNN)和卷积神经网络(Convolutional Neural Network, CNN)为代表的深度学习方法逐渐在事件提取中快速应用和发展。Nguyen<sup>[5]</sup>使用RNN来进行事件提取的研究,在神经网络输入层除了使用传统的词向量之外,还根据文本内容增加了额外的特征向量,因此,能够更好地在局部文本中提取事件。Chen等<sup>[6]</sup>提出动态多池化卷积神经网络(Dynamic Multi-Pooling Convolutional Neural Networks, DMCNN)事件提取模型,在传统CNN模型基础上增加了动态多池机制,从而提升了事件提取的效果。

基于模式匹配的事件提取由于需要进行具体模式的构建,因此,方法可移植性较差,且模式构建通常需要领域相关专家的参与。与模式匹配方法相比,基于机器学习的方法需要的人工干预相对较少,但也需要借助工具选取与任务相关的特征,因而特征选取的好坏与事件提取的效果直接关联。而基于深度学习的事件提取方法采用的是端到端的学习,不需要借助外部的自然语言处理工具设计特征,但其对话料库的质量和数量要求很高。

### 2.2 主题事件提取

单个元事件通常无法清楚描述整个事件,而主题事件作为元事件的有机组合,能更好地表现主题,目前可分为基于事件框架和基于本体两种方式。

基于框架的主题事件提取通过构建事件框架来提取事件,并根据一定规则将事件融合在一起。许荣华等<sup>[7]</sup>通过定义事件融合框架(Topic Event Fusion Framework, TEEF)来完成主题事件提取,如图 2 所示,一般通过合并与同一主题相关的所有元事件以及通过计算元事件与主题之间的相关性来呈现主题事件。赵文娟等<sup>[8]</sup>基于主题事件框架,构建网络事件提取流程,对从网络文档中提取、合并与主题事件相关的各种信息的技术和方法进行了描述与验证。

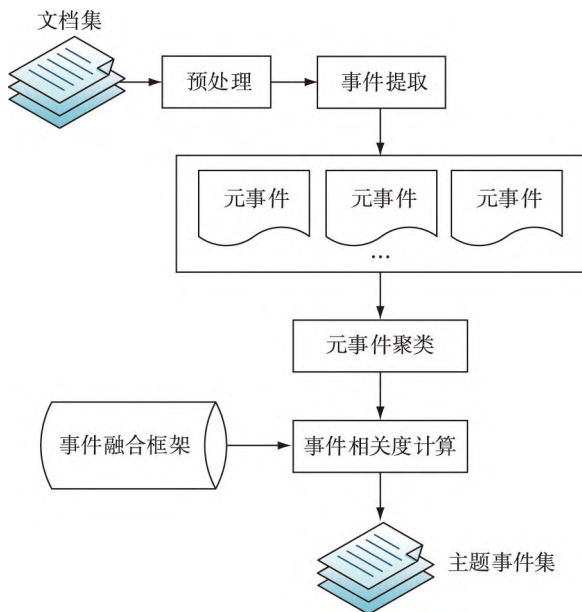


图 2 主题事件融合框架

基于本体的主题事件提取开始受到越来越多的关注。本体是概念及概念间关系的一种表示方法,可以被看作一个描述某领域知识的通用概念模型,因此,非常适合描述主题型事件。张一帆等<sup>[9]</sup>提出了事件五元组表示方式和事件本体模型,该模型是以事件类为基本单位的知识表示模型,包含了时间、地点、动作、参与者、结果等事件要素,能够更全面、准确地描述突发事件,更好地展示主题。吴奇<sup>[10]</sup>将本体技术应用到事件提取中,利用对领域知识的描述进行事件提取,指出可以利用本体中的概念和关系,结合本体中事件结构的特点,根据不同的算法和规则实现主题事件提取。

### 3 事件提取方法在军事领域的研究现状和应用分析

#### 3.1 事件提取方法在军事领域的研究现状

事件提取方法最早起源于美军对作战文书进行信息提取的需求,后来逐渐发展到金融、新闻、法律、医学等领域,并取得了极大的进步。国内近些年也开始注重对事件提取的研究,但目前,相比其他领域来说,军

事领域事件提取相关研究相对较少,仅在军事事实事件提取、战场元素建模、作战文书事件提取等方面有一些研究。

沈大川等人<sup>[11]</sup>提出了利用本体和规则推理捕获战场“关键事件”的方法,构建了战场态势核心本体以及战场领域本体,提出战场数据是以事件的形式传递的,战场“关键事件”提取规则建立在战场领域本体的基础之上,通过对战场元素的概念建模以及一定条件的约束,能反映战场的要素及要素间的基本关系,结合一定的知识和规则可将这些要素和关系聚合成“关键事件”。

宋仁亮等<sup>[12]</sup>提出利用事件描述模型提取战场关键事件的方法,他分析了战场关键事件的主要类型和相关特征,建立了关于战场关键事件的描述模型,通过对作战目标和战场区域之间的关系、作战目标与战场分界线之间的关系、战场实体属性的变化、作战力量的变化进行分析和计算,提取战场关键事件。

付雨萌等<sup>[13]</sup>以某海军舰队的活动事件为例,对相关军事事实进行了分类,在此基础上,结合军事活动的特点,分别对活动事件进行结构化、形式化描述,实现军事活动事件本体的构建,从而为其后续进行军事活动相关领域知识库及知识图谱的构建打下基础。

游飞<sup>[14]</sup>对军事装备实体事件进行分类,并运用双向长短时记忆(Bi-directional Long Short-Term Memory, Bi-LSTM)网络模型对事件触发词进行识别,通过在模型中加入负采样训练得到的特征向量,并引入句法分析和双向多层 LSTM,从而提升 LSTM 网络模型的性能,取得了良好的效果,反映出事件提取在军事领域的研究价值。

王学峰等<sup>[15]</sup>针对作战文书中出现的新力量、新编号和新战法难以通过简单的模板构建提取事件的实际问题,提出利用深度学习方法从作战文书中获取关键事件。Bi-LSTM 网络对较长句子上下文能较好记忆,动态词向量(Embedding from Language Models of Character, ELMo)对汉字语义能多重表示,CRF 模型对标注规则能有效学习,基于这些特点,构建了结合这三种方法的事件提取模型,并在演习导调文书语料集上进行了实验,取得了较好的事件提取效果。

#### 3.2 军事知识图谱的应用现状

军事知识图谱中包含的事件知识隐含于军事大数据中,需要通过事件提取技术从不断增加的海量军事数据中获取关键事件知识才能实现数据的有效利用。目前,军事领域已经逐渐开展相关知识图谱构建工作,为军事人员快速准确获取并共享军事相关知识提供支撑。



邢萌等<sup>[16]</sup>面向部队平时及战时的应用场景,针对军事领域的特点,提出军事领域知识图谱及应用技术架构,描述了军事领域知识图谱构建环节的难点,对基于本体的知识表示、基于机器学习的知识提取等关键技术进行研究,为开展军事领域知识图谱的应用提供支撑手段。

吴云超等<sup>[17]</sup>为提高仿真推演系统的效率,探讨了领域知识图谱在仿真实体动态生成中的应用。通过提出面向仿真推演的领域知识图谱构架及领域知识图谱构建方法,建立基于军事专家经验和知识的领域知识体系,实现从实时战场数据、作战条例、历史规律等结构多样的数据中提取相关的实体、关系、属性等事件要素。

车金立等<sup>[18]</sup>将知识图谱应用于装备维修保障知识库的构建。在对装备维修保障知识图谱的构建流程设计的基础上,利用装备维修保障数据进行关键信息的提取。知识图谱在装备维修保障领域的应用目的是解决装备维修保障信息化过程中出现的信息过载、查询信息效率低下等问题。

张进等<sup>[19]</sup>针对传统武器系统故障诊断方法的一些弊端,在统计岸炮武器系统各类常见故障的基础上,利用知识图谱构建领域知识库,并根据武器系统常见故障设定多重任务场景,然后基于任务驱动智能客服多轮对话,实现武器系统的故障诊断和排除。

陈辞等<sup>[20]</sup>从复杂关系语义特征出发,研究如何利用现有的军事知识图谱,对新增的军事知识进行语义融合和组织,深入研究军事语义信息提取方法以及基于在线和学习的信息提取机制,构建基于关联语义链网络的军事知识图谱演化研究方法和技术构架。

王保魁等<sup>[21-22]</sup>基于知识图谱技术,采用态势要素解析方法与形式化态势知识描述方法相结合的方式,对想定场景初始态势中实体及其关系进行分析和知识表示学习,并提出基于图嵌入的兵棋联合作战态势实体知识表示方法,为大规模联合作战态势知识的获取、融合、推理奠定基础。

胡志磊等<sup>[23]</sup>围绕以事件为核心的事件图谱,对其构建与应用的相关模型和方法进行总结。对其中包含的事件提取、事件关系推断以及事件预测等技术进行分析,并给出事件图谱具体的应用场景。

### 3.3 面向仿真推演的事件提取方法应用分析

模拟对抗演习时,推演数据急速增长,如何从这些海量、低密度、结构多样的信息中提取出关键事件逐渐受到重视。通过对推演数据进行事件提取,并以军事知识图谱、作战过程描述、作战行动脉络等形式展现,可以用来支持信息检索、自动问答、情报分析、知识推

荐等活动,辅助导演部更好地对演习进行复盘评估,实施总结讲评,从而让指挥员更清楚地了解作战要素及演习过程,更有效地总结经验教训或者实施指挥决策,具体应用框架如图3所示。



图3 面向辅助演习讲评的事件提取方法应用框架

#### 1) 作战过程分析与描述

如何从海量的计算机演习数据中筛选出影响演习进程或者作战结果的关键事件,对于分析评估整个作战过程至关重要。演习过程中,指挥员通常会根据作战任务和态势变化下达大量演习指令,形成众多的作战行动,产生不同的行动效果。这些作战行动中往往会包含一些影响战役进程或战略全局的重要事件,通过对这些重要事件进行提取和梳理不仅有助于描述作战过程,帮助指挥员聚焦关键行动,减少冗余信息的干扰,甚至可以进一步探寻联合作战中一些隐藏的规律。

#### 2) 军事知识图谱的构建

知识图谱(Knowledge Graph)是一种描述实体及其关系的语义网络,它提供对领域知识的可视化表示方法。军事知识图谱是各类作战实体及其关系的可视化呈现,通过构建军事知识图谱,将散乱、无序的战场数据整合在一起,可以提供作战要素、行动、效果以及关系等的查询与相关知识的推荐,为军事数据智能化分析提供有力支撑。事件提取作为军事知识图谱构建的基本方法之一,将对运用军事知识图谱研究作战过程中的事件及其关系建模等问题提供强有力的手段。

#### 3) 作战行动脉络分析

元事件粒度相对较小,通过其看问题往往比较片面,仅仅对其进行信息提取无法令人清晰地认识整个

事件过程。未来军事领域中,事件提取方法重要的一个应用趋势就是作战行动脉络分析。作战行动脉络作为一种特定的事件脉络,它通常是演习中指挥员关注的重点。面向演习讲评或者指挥决策需求时,导演部和指挥员更需要通过获取行动事件发生的原因、经过和结果等信息以及行动之间的层次关系、因果关系等来了解整个战役行动的脉络,掌握整个战役的发展过程,从而通过对多个行动及其之间关系的有序集合,完整、清晰地呈现整个战役的发展过程,真实再现重要作战行动的来龙去脉。

### 3.4 事件提取方法在军事领域的应用趋势

利用事件提取方法对海量异构的军事数据进行分析与挖掘,能够大幅提高军事大数据的综合利用能力。随着人工智能技术的发展,事件提取方法在军事领域中的应用将呈现以下趋势。

#### 1) 关注事件溯源以及趋势研判

现代战争,军事行动多样,作战样式复杂,涉及要素众多,如何从多源、无序、复杂的海量数据中发现关键军事事件并对其来源以及意图趋势进行分析研判,对辅助指挥员准确判断敌情、正确指挥决策极为重要。事件提取方法作为情报分析与态势研判的基础,能够为指挥员从杂乱无章的数据中发现关键军事事件,并进行因果关系研判提供基本手段。通过对事件溯源和趋势分析,对弱关联事件进行数据挖掘,可以将军事行动的整个过程以及行动间的关系清晰地展现出来,为指挥员决策提供依据。

#### 2) 聚焦特定任务知识图谱

目前,军事知识图谱已发展得较为成熟,但其建设与应用仍存在较大局限。随着军事智能化要求的提高,军事知识图谱将进一步细化到各分支的业务领域。今后,面向特定作战任务的知识图谱以及侧重于复杂事件及其关系建模的知识图谱,将越来越受到关注。因此,针对不同的军事任务,需要从实际应用角度出发,考虑具体的应用背景和知识框架,界定出合理的知识粒度,才能更好地实现面向特定任务知识图谱的事件提取。

#### 3) 重视面向事件的语料库构建

目前,高质量数据集缺失或不足问题,使事件提取在军事领域的应用仍有较大局限。尤其深度学习方法对数据质量和数量要求很高,需要大量不同实例的数据集作为训练样本数据,数据量达不到一定规模将难以开展基于深度学习的事件提取研究。而当前军事领域面向事件提取的语料库并不丰富,成为制约军事领域事件提取方法研究的瓶颈。因此,未来将重视军事领域语料库的扩建,解决相关领域语料缺乏的问题。

## 4 结束语

本文梳理了事件的概念、事件提取的发展历程,对元事件和主题事件的方法分别进行归纳和分析,并结合事件提取在军事领域研究现状和技术发展,指出了事件提取方法在军事领域今后可能的应用趋势,为下一步事件提取工作的开展和研究提供参考。

### 参考文献:

- [1] Allan J, Papka R, Lavrenko V. On-line New Event Detection and Tracking[C]//Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 1998:37-45.
- [2] Yang Y, Carbonell J G, Brown R D, et al. Learning Approaches for Detecting and Tracking News Events[J]. IEEE Intelligent Systems Special Issue on Applications of Intelligent Information Retrieval, 1999, 14(4):32-43.
- [3] 杨竣辉.文本事件关系抽取中关键技术研究及应用[D].上海:上海大学,2019.
- [4] 王伟玉,史存会,俞晓明,等.一种事件粒度的抽取式话题简短表示生成方法[J].山东大学学报,2021,56(5):66-75.
- [5] Nguyen T H, Cho K, Grishman R. Joint Event Extraction via Recurrent Neural Networks[C]//Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2016:300-309.
- [6] Chen Yubo, Xu Liheng, Liu Kang, et al. Event Extraction via Dynamic Multi-Pooling Convolutional Neural Networks[C]//Association for Computational Linguistics, 2015:167-176.
- [7] 许荣华,吴刚,李培峰,等.基于事件框架的主题事件融合研究[J].计算机应用研究,2009,26(12):4542-4544.
- [8] 赵文娟,刘忠宝.基于汉语框架的网络事件抽取及相关算法研究[J].情报理论与实践,2016,39(10):112-116.
- [9] 张一帆,郭勇,李坤伟,等.一种突发事件领域本体建模方法[J].信息系统工程,2020,1(5):134-136.
- [10] 吴奇.基于领域本体的 Web 实体事件抽取问题研究[D].济南:山东大学,2014.
- [11] 沈大川.战场关键事件提取技术研究[J].计算机技术与发展,2009,19(11):202-205,209.
- [12] 宋仁亮,戴兆乐.战场关键事件提取与告警方法[J].软件工程,2016,19(10):1-3.
- [13] 付雨萌,程瑾,罗准辰,等.基于本体的军事活动事件知识建模研究[J].中华医学图书情报杂志,2020,29(3):47-52.
- [14] 游飞.基于深度学习的军事事件抽取研究[D].南京:华

- 东计算技术研究所,2018.
- [15] 王学峰,杨若鹏,李雯.基于深度学习的作战文书事件抽取方法[J].信息工程大学学报,2019,20(5):635-640.
- [16] 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.
- [17] 吴云超,毛少杰,周芳.面向仿真推演的领域知识图谱构建技术[J].指挥信息系统与技术,2019,10(3):32-36.
- [18] 车金立,唐力伟,邓士杰,等.装备维修保障知识图谱构建方法研究[J].兵工自动化,2019,38(1):15-19.
- [19] 张进,徐宁骏,赵伟光,等.基于智能客服技术的武器系统故障诊断方法[J].指挥控制与仿真,2020,42(4):123-127.
- [20] 陈辞.基于知识图谱的军事知识演化技术研究[J].舰船电子工程,2019,39(6):22-27.
- [21] 王保魁,吴琳,胡晓峰,等.基于知识图谱的联合作战态势实体描述方法[J].指挥控制与仿真,2020,42(3):8-13.
- [22] 王保魁,吴琳,李丽,等.基于图嵌入的兵棋联合作战态势实体知识表示学习方法[J].指挥控制与仿真,2020,42(6):22-28.
- [23] 胡志磊,靳小龙,陈剑赞,等.事件图谱的构建、推理与应用[J].大数据,2021,7(3):80-96.
- (责任编辑:张培培)