

基于本体的军事知识获取和知识分析^{*}

Ontology-based Military Knowledge Acquisition and Knowledge Analysis

雷玉霞

(曲阜师范大学计算机系) (中国科学院软件研究所99研 北京100080)

Abstract Military special knowledge plays an important role in many knowledge systems such as military decision system, national language understanding, and intelligence tutor system. In this paper, we introduce an ontology-based method for acquiring and analyzing military knowledge, and we have found a set of axioms of military knowledge in order to check semantic consistency, completeness and accuracy. These axioms are also essential in reasoning and interconnecting with military knowledge.

Keywords Knowledge acquisition, Military ontology, Knowledge analysis, Military axiom

1. 引言

军事专业知识在军事决策系统、咨询系统、智能教学系统(ITS)、自然语言理解等系统中具有重要的作用。美国学者认为,复杂、实时的军事问题求解依赖于广泛的知识,为此美国军方投资创建了一个大型的知识库系统(即 HPKB),以用于未来复杂的高科技战争^[1]。显然,军事问题求解更需要足够的军事专业知识。

在建立知识库系统过程中,知识获取是最关键同时又是很困难的问题之一^[2~6]。概括地讲,有两种途径:一是从领域专家处获取专业知识,二是从文本中直接获取。由于知识工程师直接从领域专家处获取知识是一个非常复杂的人到个人的交互过程,具有很强的个性,没有统一的办法,很难得到完整的、系统的学科知识,而且据统计90%以上的知识可以从文本中获取,因此,对从事大规模知识获取的人员来说,从文本中获取知识无疑是一种更直接有效的方法。近几年,从文本中获取知识已成为一个重要的研究领域^[4,5]。但是,文本知识往往是不规范的、复杂的,自然语言理解还有很多技术难点。因此,在没有一定量的“背景知识”时,实现完全自动获取是不现实的。例如,若抽取飞机自卫电子对抗设备这一概念,一个获取算法需要一定的背景知识(如电子对抗设备)。综合分析,我们采用“人机交互”的半自动获取方法(见图1)。

知识分析是建立知识库系统的关键一环,其目的是检查知识库中的知识语义一致性、完备性和精确性。我们主要分析概念本体中的属性和关系。近年来,本体(Ontology)已成为知识工程领域很普遍的概念,在AI领域中,本体是指“概念化的明确表示”^[7]。许多大型知识系统都不同程度地应用了本体,如美国军方的大型知识库系统 HPKB,美国 D.

Lenat 教授等研制的大型常识知识库系统 Cyc,普林斯顿大学的重要研究成果—WordNet 等。

我们引进本体就是想用属性集明确刻画领域概念,用关系集刻画领域概念之间的关联,以建立一个健全、简洁、有良好联通性的军事知识库。我们的军事本体有两个层次:一是军事领域本体,是对军事领域的概念结构进行整体刻画;二是概念本体和属性本体。属性本体是基于本体的简洁性、可重用性以及知识联通等方面考虑而设计的。

本文安排是:第2部分讨论军事本体的设计思想及其本体的构件;第3部分介绍如何从《中国大百科全书》军事卷电子版和《军事辞海》中获取专业知识;第4部分讨论基于本体的知识分析方法,以进行知识库的语义一致性、完备性以及精确性检查;最后总结全文并提出需要进一步探讨的问题。

2. 军事知识本体

每个研究领域都有各自的研究对象,都涉及到有哪些概念、属性和关系;属性与属性之间、关系与关系之间有哪些约束关系;知识表示模型等重要的问题。本体可以很好地回答这些问题。在军事领域中,研究对象是武器装备、战争和军事组织等。在本体设计中,我们充分考虑了完备性、精确性、可重用性、简洁性和可扩展性等重要原则^[7]。

我们的军事知识本体分为两个层次:一是军事领域本体;二是军事概念本体和属性本体。军事领域本体描述军事领域中的概念结构以及概念本体系统。军事概念本体主要包括两个部分:属性部分和关系部分。属性集部分描述概念的内涵,关系表示与其他概念的关联。概念本体可以实现属性本体中的部分属性。属性本体只能作为属性。例如速度属性本体。

^{*}本文的工作是在中科院计算所大规模知识处理课题组完成。

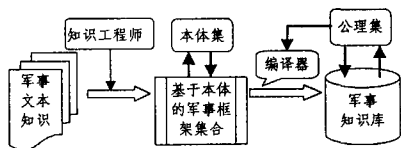


图1 基于本体的军事文本知识获取

2.1 军事领域本体

形式上,军事领域本体是个七元组: (C, R, A, T, H, O, F) 。其中 C 表示所有军事概念的集合,我们用 C 表示概念; R 表示所有关系的集合; A 表示所有属性的集合; T 不在 C 中,但在分类 H 中, T 表示所有实体的概念,是 C 中所有概念的上位概念; $H = (C, \leq)$ 表示 C 中概念的分类结构,其中 $\leq$ 是一个偏序; $(O, \leq)$ 表示所有军事概念本体集合 O 上的结构,其中 $\leq$ 是个偏序,满足:

(2.1) 对任何 $O \in O, O' \in O, O \leq O'$ 蕴含 $O \subseteq O'$; 并且

(2.2) 对任一元素 $O \in O, O = \bigcup_{O' \leq O} O'$

$F: H \rightarrow O, F$ 是一个满射, 且是连续的, 即满足:

(2.3) 对任意的 $C, C' \in H, C \leq C'$ 蕴含 $F(C) \leq F(C')$; 并且

(2.4) 对每个 $O \in O$, 至少存在一个 $C \in H$, 使得 $F(C) = O$; 并且

(2.5) 对每个 $C \in H, F(C) = \bigcup_{C' < C} F(C')$

2.2 军事概念本体和构件解释

概念 C 的本体记为 $O(C)$ 。形式上, $O(C)$ 是个三元组 $(A(C), R(C), P(C))$ 。其中, $A(C) \subseteq A, R(C) \subseteq R, P(C)$ 表示侧面集合, $A_s(C)$ 表示与 $O(C)$ 有关的公理集。概念本体通过继承、部分关系构成一个概念本体系统。概念本体可以实现属性本体中的部分属性, 任给概念本体 $O(C) \in O$, 则 $O(C)$ 满足下面三个条件:

(2.6) 如果 $C \leq C'$, 则 $O(C) \subseteq O(C')$;

(2.7) 如果 $C \in C$, 则 $O(C) \supseteq \bigcup_{C' \in C, C' < C} O(C')$, 其中 $P(C', C)$ 表示 C' 是 C 的物理部分;

(2.8) 如果 C 是 C' 的一个物质, 则 $O(C) \subseteq O(C')$ 。

2.2.1 属性和属性分析 在自然言语中, 任何概念都需从不同角度予以表达。就战争而言, 不同时期的战争都使用一定的武器, 都是在一定的时空中的进行的^[5]。诸如涉及到爆发时间、爆发原因、爆发地点、使用的武器、伤亡总人数和战略战术等方面。它是军事领域中一个核心概念。为了保证属性名的规范性和标准性, 我们参阅了《中华人民共和国军用标准》。

本体中的属性都有一个值域、类型和时变性等。对于一些特殊的属性还要考虑属性值的概率分布情

况(如正态分布)。为什么要考虑属性值的概率分布? 这 and 具体应用有关。例如在自然语言查询系统中, 若用户提问“潜艇的水上排水量一般是多少?”。我们可以把“水上排水量”看成随机变量, 采用统计技术解决类似的问题。

2.2.2 关系和侧面 MindNet 等对语义关系进行了深刻分析, 提出了24类语义关系型^[10]。同样, 我们在设计本体时, 关系是不可缺少的, 因为关系是描述领域中概念之间的关联。设 R 是一个 N 元关系, 则形式定义为 $R \subseteq C_1 \times C_2 \times \dots \times C_N$, 其中 $C_1, C_2, \dots, C_N$ 表示 N 个领域概念。概念本体中的每个关系都有类型、定义域(Domain)和值域(Value)等。例如“有部分/是部分”还是非对称的。

象“伤亡人数”这样的属性, 若不给其他限制容易产生歧义, 例如是哪一方的伤亡人数。为准确定义属性或关系, 我们引入了侧面。侧面是对属性或关系的进一步限定。例如参战方、方式、程度一范围和可信度等都是侧面。对于公理, 我们放在知识分析中讨论。

2.2.3 概念本体实例

Defontology 战争本体 继承 基础本体

```
{
  属性:汉语拼音
    :类型 字符串
  属性:英文
    :类型 字符串
  属性:爆发时间
    :类型 字符串
    :同义词 开始时间
    :时变性 连续变化
  属性:结束时间
    :类型 字符串
    :时变性 连续变化
  属性:参战一方
    :类型 字符串数组
  属性:参战兵力
    :类型 字符串数组
    :侧面 参战方
  属性:作战过程
    :类型 过程本体
    :同义词 作战经过
  .....
  属性:伤亡人数
    :类型 整数
    :单位 个
    :侧面 参战一方
  属性:被俘人数
    :类型 整数
    :单位 个
    :侧面 参战一方
  属性:伤亡与被俘人数
    :类型 整数
    :单位 个
    :侧面 参战一方
  属性:伤亡总人数
    :类型 整数
    :单位 个
  属性:伤亡与被俘总人数
    :类型 数量
    :单位 个
  关系:是
    :类型 字符串数组
    :注释“关于战争的评价”
  ..... }
```

图2 战争本体(部分)

3. 基于本体的军事知识获取

知识库是机器智能的重要来源。在建立知识库时,首要的问题是要选好知识源。一个好的知识源应该是正确的、规范的、科学的和权威的,并且能涉及到军事领域中的各个方面。基于以上考虑,我们采用《中国大百科全书》军事卷电子版和《军事辞海》作为我们的知识源。

我们的知识获取方法主要有三步:第一步“人机交互”的翻译阶段,即将军事文本翻译成框架集合;第二步是语法检查阶段。编译器对单个框架进行检查。若有语法错误,进行提示;若符合语法,则进行编译;第三步编译器将相关的知识进行合并、连接。

知识获取的翻译阶段是必要的。首先,在自然语言理解方面,虽然已经基本上掌握了单个句子的理解技术,但是还很覆盖全面的语言现象,特别是对于整个段落或篇章的理解还有很多技术难点;其次,不同知识源中的知识有时存在一致性、描述粒度和精度不同。所以要保证所获取的知识是正确的、精确的和完备的,人工干预是必要的。下面给出了索姆河会战和百团大战的框架表示(见图3)。

```
Defframe 百团大战:战争本体
{
  汉语拼音: Baituan Dazhan
  英文: Great Campaign with One Hundred Regiments
  开始时间: 1940年8月20日
  结束时间: 1940年12月5日
  :注释“战役基本结束时间”
  交战一方: 中国八路军
  交战另一方: 日军, 和中国伪军
  持续时间: 3个半月
  伤亡人数: 大于20000人
    :参战方 日军
  伤亡人数: 大于5000人
    :参战方 中国伪军
  被俘人数: 大于280人
    :参战方 日军
  被俘人数: 大于18000人
    :参战方 中国伪军
  伤亡人数: 大于17000人
    :参战方 中国八路军
  .....
}
Defframe 索姆河会战:战争本体
{
  汉语拼音: Suomuhe Huizhan
  英文: Somme Battle of
  爆发时间: 1916年6月24日
  结束时间: 1916年11月中旬
  战场: 法国北部索姆河地区
  交战一方: 英军, 和法军
  交战另一方: 德军
  伤亡人数: 大于450000人
    :参战方 英军
  伤亡人数: 大于340000人
    :参战方 法军
  伤亡人数: 538000人
    :参战方 德军
  伤亡总人数: 约1340000人
  是: 第一次世界大战中规模最大的战役作战过程: 索姆河会战的经过
  :注释“事件序列”
  .....
}
```

图3 百团大战和索姆河会战的基于战争本体的框架表示

4. 基于本体的军事知识分析

知识分析是建立知识库的关键一环,其目的就是帮助检查知识库中的知识语义一致性、完备性和精确性。为此,我们设计出一个基于概念本体的公理(主要是用一阶逻辑来描述)系统。我们的知识分析主要有两点:一是本体中的属性分析;二是本体中的关系分析。主要分析属性的值域,时变性以及属性之间的约束关系,分析关系本身的基本性质以及关系之间的关联。

4.1 本体中的属性分析

本体中的属性之间往往具有一些约束关系。我们以战争本体为例给出一些公理(注意,W要有具体的实例代替),以说明我们设计公理的思想。设 $X \in \{\text{伤亡, 死亡, 失踪, 被俘, 伤亡与被俘, 伤亡与失踪}\}$ 。注意,相似的公理进行了合并。

公理 4.1 $\forall W$: 战争本体(早于(爆发时间(W), 结束时间(W)))。

公理 4.2 $\forall W$: 战争本体(伤亡与被俘人数(参战方(W)) = 伤亡人数(参战方(W)) + 被俘人数(参战方(W)))。

公理 4.3 $\forall W$: 战争本体(伤亡与失踪人数(参战方(W)) = 伤亡人数(参战方(W)) + 失踪人数(参战方(W)))。

公理 4.4 $\forall W$: 战争本体(参战总人数(W) = 参战人数(参战方1(W)) + 参战人数(参战方2(W)))。

公理 4.5 $\forall W$: 战争本体($(X \text{ 总人数}(W) = X \text{ 人数}(\text{参战方1}(W)) + X \text{ 人数}(\text{参战方2}(W)))$)。

公理 4.6 $\forall W$: 战争本体($(\text{参战总人数}(W) > = X \text{ 总人数}(W))$)。

公理 4.7 $\forall W$: 战争本体($(X \text{ 总人数}(W) > = X \text{ 人数}(\text{参战方}(W)))$)。

公理 4.8 $\forall W$: 战争本体($(\text{参战总人数}(W) > = \text{参战人数}(\text{参战方}(W)))$)。

公理 4.9 $\forall W$: 战争本体($(\text{参战人数}(\text{参战方}(W)) > = X \text{ 人数}(\text{参战方}(W)))$)。

公理 4.10 $\forall W$: 战争本体($(Y \text{ 人数}(\text{参战方}(W)) > = Z \text{ 人数}(\text{参战方}(W)))$), 其中 $Y \in \{\text{伤亡与被俘, 伤亡与失踪}\}$, $Z \in \{\text{伤亡, 被俘, 失踪}\}$ 。

设F是个具体战争框架或战役框架。F中属性的属性值往往是不精确的(例如索姆河会战中法军伤亡大于340000人,但是可以接受的。为了描述这样的情况,我们给出几个定义,以帮助检查知识一致性、精确性和完备性。

定义 4.1(时间不一致) 称F是时间不一致的,若 $F \vdash \text{早于}(\text{爆发时间}(W), \text{结束时间}(W))$ 。

定义 4.2(人数精确一致的) 称F是人数精确一致的,若F满足公理4.2至公理4.10。

定义4.3(人数不精确一致的) 称F是人数不精确一致的,若F满足公理4.6~4.10,且有:

(4.1)公理4.2~4.4中至少有一条件不成立,或者

(4.2) $F \vdash (X \text{ 总人数}(W) \neq X \text{ 人数}(\text{参战方1}(W)) + X \text{ 人数}(\text{参战方2}(W)))$.其中X或是参战,或伤亡或死亡或失踪或被俘或伤亡与被俘。

定义4.5(人数不一致) 称F是人数不一致的,若公理4.6~4.10中有个子公理不成立。

定义4.6(人数不完备的) 如果“X人数(参战方(W))=Void”或者“X总人数(W)=Void”,其中 $X \in \{\text{参战, 伤亡, 死亡, 失踪, 被俘, 伤亡与被俘, 伤亡与失踪}\}$ 。

4.2 概念本体中的关系分析

我们以部分关系为例,进行关系分析。Winston, Chaffin 和 Herrmann(WCH)等根据功能、可分离性和同质性将部分关系区分为 Component/Integral-object, Member/Collection, Place/Area 和 Feature/Activity 等六类^[10],见表1。表中的+表示满足所对应的性质,-表示不满足。

表1 WCH的部分关系的分类

类型	功能	可分离性	同质性	举例
Component/Integral-object	+	+	-	车轮/汽车
Member/Collection	-	+	-	美国/北约组织
Portion/Mass	-	+	+	西瓜瓣/西瓜
Stuff/Object	-	-	-	钢/自行车
Feature/Activity	+	-	-	抓/堆积
Place/Area	-	-	+	绿洲/沙漠

在军事中,部分关系是最基本的关系之一。军事对象大多数是可拆分的,下面就以战斗舰艇本体为例,根据固定程度、功能和可分离性将部分关系作以下分类:

是组成/有组成:类似于WCH的Component/Integral-object,描述具体对象的物理结构关系,是固定的、可分离的并有特定的功能。例如“发动机/舰艇”。

是设备/有设备:为了实现特定的目的而增加的辅助部分,是不固定的、可分离的并有特定的用途。例如“导航设备/舰艇”。

是成份/有成份:类似于WCH的Stuff/Object,描述对象的制造元素或成份,是不可分离的,也没有明显的功能和用途。例如“钢材/舰艇”。

是结构/有结构:类似于WCH的Place/Area,描述空间关系,是固定的、不可分离的并有特定的用途。另外,这样的部分本身是一定的空间,例如“弹舱

/战斗舰艇”和“射击孔/坦克”。

是武器/有武器:具有特定的军事用途,是不固定的、分离的,但不具有任何功能的角色。例如“舰载导弹/舰艇”。

设 $X \in \{\text{有组成, 有设备, 有成份, 有结构, 有武器}\}$,则有公理:

公理4.11 $\forall u, x(X(u, x) \rightarrow \text{是部分}(u, x))$ 。

定义4.6(部分关系不一致) 设F是个框架,称F是部分关系不一致的,如果 $F \vdash X(u, x)$,并且 $F \vdash \neg \text{是部分}(u, x)$ 。

定义4.7(框架语义不一致的) 设 F_1 和 F_2 表示框架,称 F_1 和 F_2 是语义不一致的,如果 $F_1 \vdash X(u, x)$,并且 $F_2 \vdash X(x, v)$,并且 $F_1 \wedge F_2 \vdash \neg (u, v)$ 。

当然,关系部分之间的公理还有很多。我们给出这些以说明我们进行关系分析的思路。

总结和下一步讨论的问题 本文主要讨论了基于本体的军事知识获取和分析方法。文章首先介绍了军事领域本体以及概念本体的设计思想。在第3节,我们提出了基于本体的半自动知识获取方法。介绍了如何从《中国大百科全书》军事卷电子版和《军事辞海》中获取专业知识。第4部分指出知识分析是建立知识库过程中的关键一环。其目的是帮助检查知识库中知识的语义一致性和完备性,文章主要讨论了属性之间、关系以及关系之间的公理。

实践证明,我们基于本体的知识获取方法是可行的、有效的。目前,我们的知识分析是基于概念本体的,其公理仅适用于一定的子领域。将来,我们在概念本体中公理的基础上抽取更一般的公理。这是一件很有意义的问题,因为可以实现跨学科的联通。

参考文献

- Alphatech, et al. HPKB Course of Action Challenge Problem Specification. Burlington, MA. 1998
- Lei Yuxia, Cao Cungen, Sui Yuefei. Acquiring Military Knowledge from Texts in the Electronic Encyclopedia of China. ICYCS'2001, 1:367~371
- 曹存根.面向专家的知识获取.北京:科学出版社,1998
- Bowden P R, Halstead P, Rose T G. Extracting Conceptual Knowledge from Text Using Explicit Relation Markers. In: Shadbolt, N., Ohara, K, Schreiber, G. eds. Advances in Knowledge Acquisition. Lecture Notes In Artificial Intelligence, vol. 1076. Springer-Verlag, Berlin, 1996. 147~162
- Hull R, Gomez F. Automatic Acquisition of Biographic Knowledge from Encyclopedic Texts. International Journal of Expert Systems with Applications. 1999. 261~270
- Mizoguchi R. Knowledge acquisition and ontology. In: Proc. of the KB&KS'93. 1993. 121~128
- Gruber, Thomas R. A translation approach to portable ontology specifications. Knowledge Acquisition, 1993.5:199~220
- Gómez-Pérez A. Evaluation of Taxonomic Knowledge in Ontologies and Knowledge Bases
- Bowman M, et al. Ontology Development for Military Applications. In: Proc. of the 39th Annual ACM Southeast Conf. 2001
- Artale A, franconi E, et al. Part-whole relations in object-centered systems: an overview. Data&Knowledge Engineering, 1996