

基于描述逻辑的事例推理综述

孙晋永^{1,2} 古天龙² 常亮²

(西安电子科技大学计算机学院 西安 710071)¹

(桂林电子科技大学广西可信软件重点实验室 桂林 541004)²

摘要 基于描述逻辑的事例推理(CBR)是当前 CBR 研究的热点之一。首先介绍了 CBR 的起源,然后回顾了基于描述逻辑的 CBR 的发展历史,接着从 4 个方面:事例表示与组织、事例检索、事例修正和事例库维护综述了基于描述逻辑的 CBR 的研究工作,最后指出了基于描述逻辑的 CBR 目前存在的问题并相应地提出了未来的研究方向。

关键词 基于事例推理,描述逻辑,事例表示,事例检索,事例修正

中图分类号 TP18 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2014.11.001

Survey of Case-based Reasoning Based on Description Logics

SUN Jin-yong^{1,2} GU Tian-long² CHANG Liang²

(School of Computer Science and Technology, Xidian University, Xi'an 710071, China)¹

(Guangxi Key Laboratory of Trusted Software, Guilin University of Electronic Technology, Guilin 541004, China)²

Abstract Case-based Reasoning based on description logic (DL-based CBR) is one of the hot research topics in CBR. Firstly, the origin of CBR was introduced. Next, the development history of DL-based CBR was reviewed. Then research works on DL-based CBR were investigated comprehensively from four aspects: case representation and organization, case retrieval, case revision and case base maintenance. Finally, some existing problems were pointed out and correspondingly some future research directions were proposed.

Keywords Case-based Reasoning(CBR), Description logic(DL), Case representation, Case retrieval, Case revision

1 引言

事例推理,即基于事例推理(Case-based Reasoning, CBR)是 20 世纪 70 年代末发展起来的一项人工智能技术,源于认知科学中记忆在人类推理活动中所扮演的角色。它的基本思想是充分利用过去的经验来解决当前遇到的相似问题。CBR 基于如下两条基本原则^[1]: (1)现实世界是有规律的,相似的问题有相似的解决办法;(2)同类的问题会再发生。一个完整的 CBR 系统包括 4 个循环过程^[2]:事例检索(Retrieve)、事例重用(Reuse)、事例修正(Revise)和事例保存(Retain)。1977 年 Schank 等^[3]在这方面的工作被认为是 CBR 的起源。随后 Schank^[4]又探索了记忆中的事例在问题解决和学习中的作用,于 1982 年在《Dynamic Memory》一书中提出了 CBR 的认知模型和框架。在此基础上, Kolodner 等^[5]于 1983 年开发了第一个 CBR 系统——CYRUS 系统。在 80 年代后期,美国的 DARPA 项目导致了 Remind 系统的产生。这被认为是 CBR 从认知科学向人工智能方向转变的标志。此后, CBR 成了人工智能科学的一个重要研究领域。

从 CBR 被提出之后,很多知名学者和学术团体致力于

CBR 的理论和应用研究,已取得了很大进展。自 1987 年以来,国际研究界每年举行 CBR 研讨会(如 ECCBR、ICCBR)。一批优秀研究成果的发表积极促进了 CBR 的发展。迄今为止, CBR 已在智能规划、智能决策、产品设计、故障检测、模式分类、电子商务、智能教学、软件复用、法律推理、信息服务、医疗诊断、电子商务、Web 服务、机器人设计和实时游戏设计等方面得到了应用。

尽管 CBR 以认知科学为基础并以人工智能等技术予以实现,但它缺乏理论方面的基础。由于人们的实践经验,没有人会否定 CBR 的确有用,但很难在理论上严格的证明。20 世纪 90 年代,研究学者开始尝试将命题逻辑、谓词逻辑和描述逻辑应用于 CBR 中,试图将它们作为 CBR 的理论基础。基于描述逻辑的 CBR 的思想是使用描述逻辑概念来描述和索引 CBR 事例和涉及到的领域知识,在描述逻辑的推理能力的帮助下完成事例检索、事例重用、事例修正和事例库维护。描述逻辑刻画 CBR 事例知识内部结构的优势及其良好的推理能力,尤其是从知识库中显式包含的知识推导出隐含包含的知识的能力,对提高 CBR 的事例表示和事例检索的准确性以及完备性提供了很大帮助。Koehler、Kamp 和 Gomez-Al-

到稿日期:2013-12-31 返修日期:2014-05-12 本文受国家自然科学基金(60963010, 60903079, 61262030, 61363030), 广西自然科学基金(2012GXNSFBA053169)资助。

孙晋永(1978—),男,博士生,CCF 会员,主要研究方向为知识表示与推理、CBR 推理, E-mail: sunjy@guet.edu.cn; 古天龙(1964—),男,教授,博士生导师,CCF 会员,主要研究方向为形式化方法、符号计算等; 常亮(1980—),男,博士,教授,CCF 会员,主要研究方向为知识表示与推理、智能规划、形式化方法。

barran 等著名学者的研究成果奠定了基于描述逻辑的 CBR 的理论基础。

本文关注了近 20 多年来基于描述逻辑的 CBR 的研究进展,考察了基于描述逻辑的 CBR 的研究历史,并围绕基于描述逻辑的 CBR 事例表示、事例组织、事例检索、事例修正以及事例库维护对已有的研究成果进行综述,以期对基于描述逻辑的 CBR 的研究起到一定的推动作用。

2 基于描述逻辑的 CBR 的发展历史

2.1 基于描述逻辑的 CBR 简介

描述逻辑^[6](Description Logic, DL)又称为术语逻辑,是基于对象的知识表示形式化方法,吸收了 KL-ONE 的主要思想。它的优点在于:具有清晰的模型理论语义;对概念性知识的处理,特别是对概念分层的处理非常有效;提供了有效的推理服务,实现了知识表达能力和推理可判定性的统一。描述逻辑通过定义领域的概念及其结构关系(角色)来刻画领域内的个体信息。在概念和角色的描述之上,其由构子从简单概念和角色构造出复杂概念和角色。描述逻辑的知识库由概念公理集合(TBox)和个体断言公理集合(ABox)两部分组成。其中 TBox 是概念描述及其关系的集合,包含了概念定义和概念的包含关系。ABox 是描述个体和关系断言的集合。ALC 是最基本的一种描述逻辑语言。在 ALC 的基础上,增加数量约束、函数性约束、定性数量约束、模糊语义约束等构子分别演变为 ALCN、ALCF、ALCQ 和 f-ALC。描述逻辑在语义 Web、数据模型、知识工程和软件工程等领域已得到了应用。

在基于命题逻辑和谓词逻辑的 CBR^[7,8]的研究思想的启发下,一些学者将描述逻辑引入 CBR 中。基于描述逻辑的 CBR 使用描述逻辑概念来描述一类 CBR 事例,并以此作为这类 CBR 事例的索引。使用知识库中的 ABox 表示具体的事例,使用 TBox 表示事例结构和涉及到的领域知识。基于描述逻辑的推理服务如概念分类操作、实例检测、可满足性检测等完成事例检索、事例重用、事例修正和事例库维护。

2.2 基于描述逻辑的 CBR 的发展历程

基于描述逻辑的 CBR 自提出以来已有 20 多年的研究历史,完成了从方法探索到方法完善的发展过程。下面按时间顺序介绍基于描述逻辑的 CBR 研究的发展历史。

早期的研究者致力于提出基于描述逻辑的 CBR 可行的实现方法。最早把描述逻辑用于 CBR 的学者是 Beck H. W. 1991 年,Beck^[9]将描述逻辑 CANDIDE(派生自 KANDOR)应用于 CBR 系统,借助概念包含关系来评价事例间的相似性。推理涉及两个过程:推演过程和归纳过程。其中推演过程用于两个事例间存在精确匹配的理想解的情况;归纳过程用于当两个事例间不存在精确匹配解时,发现和评价事例间的部分匹配并进行推理。Beck 还提出了一种异常事例,但只进行了操作上的说明,并没有给出形式化的描述。1994 年, Ashley 等^[10]用框架系统 LOOM 及其一阶逻辑语言表示法律事例间的相关性准则,以此支持事例检索,并建立了 CATO 系统来进行验证。Ashley 等认为形式化逻辑是表达 CBR 事例间的相关性准则的有效且自然的方法,并乐观地指出其它形式化逻辑也适用于 CBR。

1994 年和 1995 年, Koehler^[11,12]将描述逻辑 ALC 用于

基于事例的规划系统 MRL 中的规划检索,并在 Unix 的邮件域中应用。Koehler 给出了描述逻辑和规划逻辑结合的知识表示法,以 ALC 作为规划事例库的查询和检索语言。事例由 ALC 概念描述并索引。Koehler 的研究虽然尚待完善,但为基于描述逻辑的 CBR 研究指明了方向。

鉴于 Ashley 等^[10]和 Koehler^[11,12]使用的描述逻辑只能描述抽象域的知识,1996 年 Kamp^[13]用描述逻辑 C_{TL} (能够描述数值、串和符号集等具体域信息)实现了一个用于故障诊断的二级支持 CBR 系统。1998 年, Salotti 等^[14]使用描述逻辑 C-CLASSIC 实现了司法裁判系统的事例检索, Coupey 等^[15]研究了目标事例与历史事例的部分匹配问题,将描述逻辑 ALN_{δ} (包含默认联接词 δ 和异常联接词 ϵ)应用于法国电话网络的故障诊断与恢复中。Coupey 等指出描述逻辑的结构化事例描述能力为事例表示及索引提供了一种灵活的、富有表现力的方式。描述逻辑的自动概念分类和实例识别能力为事例组织提供了有力的支持,而且它的声明式语义有助于引入领域知识。

1999 年, Gomez-Albarran 等^[16]总结了前人的研究成果,提出了基于描述逻辑的 CBR 系统开发模型。该模型的思路是使用描述逻辑形式化地表示 CBR 系统的结构化知识,基于描述逻辑的推理机制实现 CBR 系统中的事例检索、事例修正与学习等。该开发模型是对 Koehler 的研究理论的发展与完善,为进一步开展基于描述逻辑的 CBR 的研究奠定了坚实的理论基础。此后的研究者基本上都采用了这种开发模型。

到 20 世纪 90 年代描述逻辑已经用于多个领域的 CBR 系统中,如法律(Ashley 和 Aleven 1994)、规划(Koehler 1994; Napoli 等 1996)、服务支持(Kamp 1994; Kamp 等 1996)、软件重用(Fernandez-Chamizo 等 1996)、自传数据的检索(Kamp 1997)等^[17]。

进入 21 世纪以后,研究者致力于关键问题的解决方法的完善,如更准确的事例表示方法、更高效的事例检索方法和事例修正算法,以期提高 CBR 系统的推理效率。研究者认识到早期用于 CBR 的描述逻辑的表达能力不足以表示复杂的 CBR 事例。于是一些表达能力更强的描述逻辑如 ALC-NR^[18]、SHI^[19]等被陆续引入到 CBR 中。作为事例检索中的关键环节,事例间的相似性度量自然成为这个阶段的研究热点之一。针对不同的描述逻辑语言表示的事例,研究者提出了多种相似性度量方法。2005 年 d'Amato 等^[20]、2006 年 Janowicz^[18]、2009 年 Janowicz 等^[19]分别给出了描述逻辑 ALN、ALCNR、SHI 下的事例相似性度量方法。

2010 年和 2011 年, Cojan 等^[21,22]将描述逻辑 ALC 用于 CBR 系统中的事例修正,并给出了使用扩展的 Tableau 方法、结合领域知识的事例修正算法。该算法对基于描述逻辑的事例修正算法的后续研究具有重要的指导意义。

1990 年, Aamodt^[23]提出知识密集型 CBR(Knowledge Intensive Case-Based Reasoning, KI-CBR),该推理依赖于领域知识的一个形式化模型。进入 21 世纪以来, KI-CBR 逐渐成为研究的热点之一。2004 年, Aamodt^[24]使用领域本体来建模 KI-CBR 中的通用领域知识。2007 年, Recio-Garfa 等^[25]使用 jCOLIBRI 架构开发了一个基于本体的 KI-CBR 应用。其中使用本体建模通用领域知识,基于描述逻辑的推理能力实现 CBR 推理,提高了推理的准确性和可靠性。同年, Juarez

等^[26]使用基于描述逻辑的 OWL DL 本体表示医疗领域的事例来进行医疗诊断推理。表达能力更强的描述逻辑可以更准确地表示事例,为事例检索的准确性和完备性提供了更大的支持。为了解决资源及领域知识分散的问题,研究者提出了分布式 CBR 系统^[27]。2005 年和 2013 年, d' Aquin 等^[28,29]研究了多本体环境中的分散式 CBR (Decentralized Case-based Reasoning, DzCBR),并结合语义 Web 技术实现了一个医疗决策系统 KASIMIR。该决策系统为癌症医疗诊断提供了准确的支持。该系统基于分布式描述逻辑 (Distributed Description Logics),使用由 C-OWL 实现的上下文本体来表示领域知识和修正知识。每个上下文本体代表一个特定的视角,包含解决本地问题所必需的知识。整合多个视角的本地问题解答就得到了目标问题的解答。

3 基于描述逻辑的事例表示与事例组织

3.1 基于描述逻辑的事例表示

事例表示、事例组织与事例库维护是 CBR 研究的首要问题^[11,12]。从问题求解的角度, CBR 事例表示应该包含问题描述和解描述,可以用有序对〈问题描述、解描述〉或三元组〈问题描述、解描述、解评价〉来表示。其中问题描述是事例表示中最重要的部分,如何准确地表示事例的问题描述部分是研究学者主要关注的问题。

借鉴人工智能中的知识表示方法,目前已建立了多种事例表示方法,主要有:特征向量表示法、结构化表示法、文本表示法和专门应用表示法等。特征向量表示法用一组属性-值对组成的向量表示事例,可以准确地表示事例的各属性的值,其缺点是不能描述事例的内部结构及事例之间的关系。结构化表示法可以实现事例内部结构(如层级结构、网络结构等)的描述。使用框架是该方法的一种传统形式,这种表示法已经用逻辑实现了部分形式化^[30]。

Koehler^[11,12]首次提出了基于描述逻辑 ALC 的事例表示方法:从问题描述中抽象出事例的特征,用 ALC 概念描述这些特征并以此来索引事例。基于描述逻辑的概念包含公理,索引的分层结构形成了一个有向无环图。这种事例表示方法可以更细致地刻画事例的内部结构,对提高事例表示与事例检索的准确性以及事例检索的完备性有很大帮助。此后的研究者都采用了 Koehler 的事例表示方法。

在文献^[24]中,领域本体被用来建模 KI-CBR 中的领域知识。将本体应用于 CBR 中的优势之一是可以为事例和查询描述提供统一的词汇表。使用基于描述逻辑刻画的 本体,可以保证事例描述和查询描述形式上的统一,以便事例的表示、组织和检索。文献^[25,26,28,29]使用与本体等价的描述逻辑概念来索引事例,为事例描述和推理提供了一个高效的形式化方法。

然而目前用于 CBR 事例表示的描述逻辑未能引入一些实际应用领域常见且重要的约束或语义(如数量约束、有型域约束、时序语义、模糊语义等),从而不能准确地表示具有这些约束需求的复杂事例。这在一定程度上降低了事例检索的准确性,也必然影响了 CBR 系统的推理效率。因此,根据 CBR 的实际应用领域的需要,将具有数量、有型域、时序或模糊语义等约束能力的描述逻辑引入 CBR 中以提高事例的表示能力是十分必要的。

3.2 基于描述逻辑的事例组织

CBR 事例库的设计与组织关系到事例检索的速度和成功率。为了提高事例检索的效率,必须对事例库的组织方式进行设计,特别是大规模事例库。简单的线性组织结构(如表、队列)在事例检索时效率低下。传统 CBR 系统大部分采用 Schank 的动态存储模型或 Proters 的以语义网为基础的存储模型,或者是二者的变形^[31]。这几种存储模型都是一种等级制的网状结构,其核心思想都是按一定的规则对事例进行分类并建立索引目录。这种方法对事例分类的准确率要求较高。

在基于描述逻辑的事例表示下,事例库的组织方式需要随之调整。Koehler^[11,12]使用描述逻辑概念作为事例的索引,借助概念分类操作,将事例库构建为一个包含层次组成的格结构或索引树。Salotti 等^[14]、Coupey 等^[15]、Gomez-Albarran 等^[16]也采用了这种观点。

领域本体的提出为事例库组织提供了良好的结构,这一点在文献^[25,26,28,29]中有深入的研究。本体的分类层次结构与基于描述逻辑的格结构或索引树相匹配,使得基于描述逻辑的事例库可以很自然地组织成本体的形式。

在基于描述逻辑的 CBR 系统中,事例库对应于描述逻辑的知识库。通过使用 TBox 表示事例结构定义和领域通用知识,使用 ABox 表示具体事例,将事例和领域知识统一组织在本体中,形成了一致的事例组织模型^[32]。如图 1 所示,图中 TBox 包含了事例的分类及包含关系等知识,为事例的索引。

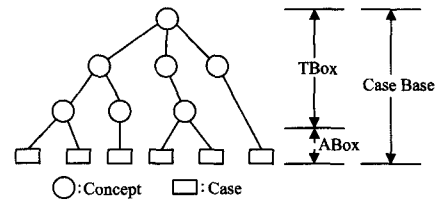


图 1 基于描述逻辑的简单事例库组织结构

在简单事例表示基础上,增加 TBox 和 ABox 中的知识描述可以实现复杂事例的表示和组织。相比 Schank 的动态存储模型和 Proters 的存储模型,其将事例定义成事例实例、事例结构和领域知识这 3 个部分并统一在本体中。这样可以实现事例知识与领域知识的融合,提高了 CBR 使用领域知识的能力,解决了传统 CBR 系统中事例知识与领域知识不易共享的问题,也使得事例结构的扩展和事例库的维护更加简单^[32]。

4 基于描述逻辑的事例检索

事例检索是 CBR 系统的核心。在事例检索前,需要对目标问题的描述进行分析、识别和提取相关特征,得到目标问题的事例表示形式,即目标事例。为了方便表述,事例库中的历史事例常被称为源事例。事例检索的主要任务是从事例库中检索出与目标事例最相似的源事例。事例检索的准确性与否很大程度上决定了 CBR 系统的推理效果。

4.1 基于描述逻辑的事例检索方法

在基于描述逻辑的 CBR 研究中,事例检索是近 10 年来研究者关注的重点。目前已经提出了多种基于描述逻辑的推理能力的事例检索方法。

Koehler^[11,12]提出的事例检索方法如下:先使用描述逻辑

概念描述目标事例,得到其索引概念,即所要执行的查询;然后使用概念分类操作将这个索引概念存储到事例库(即索引库)中。索引库对应着一个有向无环图,分类的结果决定了目标事例的索引在索引库(对应着事例库)中的位置。最后根据目标事例索引到源事例索引之间的路径长度进行排序,具有最短路径的源事例索引就是最相似事例的索引。最相似事例索引的所有实例就是候选事例。从事例库中可能会检索出多个候选事例,因此需要一个评价指标来对这些候选事例进行排序,以找出最好的一个。该指标是根据对候选事例进行修正以满足目标事例所需付出的努力的大小来进行评价的,所需付出努力最小的就是最好的候选事例。Koehler 的检索方法的缺点是:如果目标事例与源事例之间没有包含关系,事例检索过程就会失败。

Gomez-Albarran 等^[16]提出的 CBR 系统开发模型将事例检索分两个过程:首先从事例库中检索出一组候选事例,称为事例过滤操作;然后使用一个相似性函数对它们进行评价。实现事例过滤操作有两种方法:基于概念分类的检索、基于实例识别的检索。(1)基于概念分类^[33]的检索方法与 Koehler^[11,12]的事例检索方法相似。(2)基于实例识别的检索,先根据目标事例的描述特征创建一个事例实例和相应的断言,再使用实例识别操作来检索这个事例实例的最具体概念 MSC(Most Specific Concept)的所有实例,这些实例就是候选事例。最后,依据它们与目标事例的相似性来进行评价,选出最相似性的候选事例。在事例检索过程中,可以定义一个相似性阈值,只有与目标事例的相似性高于这个阈值的事例实例才被检索出来。该模型还允许在计算两个事例的相似性时结合领域知识。Gomez-Albarran 的算法成为以后事例检索算法研究的指导思想。

Kamp^[13]使用能够描述具体域信息的描述逻辑 C_{TL} 实现一个故障诊断系统。基于描述逻辑的基本推理操作,Kamp 提出了 3 种事例检索方法:基于检索的方法、基于概念分类的方法和基于对象分类的方法。其思想与 Gomez-Albarran^[16]的事例检索思想是一致的。Juarez 等^[20]提出了基于描述逻辑本体的事例检索准则,也采用了 Gomez-Albarran^[16]的事例检索思想。

Koehler^[11,12]、Kamp^[13]只考虑了目标事例与源事例精确匹配的情况,Coupey 等^[15]研究了目标事例与源事例部分匹配的问题。基于描述逻辑的概念分类操作和继承关系进行事例检索,并使用优先准则来选出一个最合适的事例。检索过程分为两个层次:概念层和实例层。假设符号 C_{new} 表示目标事例的索引。(1)在概念层中,使用分类操作在事例库中检索与 C_{new} 相似的事例索引,并且 C_{new} 与这些索引有继承关系。为了找到 C_{new} 与之有继承关系的其它索引概念,通常增加异常操作来细化 C_{new} 。该操作假定 C_{new} 针对于特定的默知识是意外情况,这样就可以处理部分匹配的问题。(2)在实例层中,计算目标事例索引和相似事例索引的最小公共包含概念 LCS,然后分析这些 LCS 概念之间的继承关系。最具体的 LCS 概念对应的事例被认为是与目标事例最相似的事例实例。

在基于本体的 CBR 中,本体为用于事例相似性评估的语义推理方法提供了知识源^[29]。Recio-Garíá 等^[25]提出的基于本体的事例检索方法以及 d' Aquin 等^[28,29]提出的分散式

CBR 系统中的本地事例检索方法均基于 Gomez-Albarran^[16]的思想。

由以上得出,目标事例与源事例的相似性评估是事例检索的关键步骤。合适的相似性度量方法有助于在候选事例集中迅速、准确地查找到最相似事例。下一节为事例相似性度量研究的现状。

4.2 事例的相似性度量方法研究

到 20 世纪 90 年代,基于描述逻辑的事例检索的基本方法已经成熟,主要以 Koehler^[11,12]、Gomez-Albarran^[16]的方法为指导思想。随着表达能力更强的描述逻辑被引入到 CBR 中,事例相似性度量逐渐成为研究热点。相比于传统的相似性度量方法,基于描述逻辑的相似性度量可以充分利用描述逻辑的表示和推理能力,可以更贴切地度量事例间的相似性。目前,基于描述逻辑的事例相似性度量方法主要有数值型度量方法和符号型度量方法两种。

4.2.1 数值型度量方法

该方法使用度量公式来评估事例索引概念间的相似性,相似性被归一化为 $[0,1]$ 内的数值。按采用的计算方法来分,其可以分为两类:基于概念外延的相似性度量和基于概念内涵的相似性度量。

(1) 基于概念外延的相似性度量

基于外延的方法最初来源于 Jaccard 的相似性度量方法^[34]和 Tversky 的对比模型^[35]。该度量方法为两个待比较的事例索引概念指定一个正比于它们外延重叠的相似性值。目前已经提出了多种基于外延的相似性度量方法。代表性的方法有 3 种:基于信息内容模型的相似性度量^[36]、基于特征模型的相似性度量^[36]和基于集合理论的相似性度量^[37]。这类方法的主要问题是没考虑概念在层次体系中的隐含语义,因此它们不能准确地度量语义相近但没有公共实例的概念间的相似性,即不能识别出不相交的事例索引概念间的相似性。

(2) 基于内涵的相似性度量

基于内涵的方法利用事例索引概念的结构来度量它们的相似性,其主要有基于语义网络模型的相似性度量^[36,38]和基于描述逻辑范式的相似性度量^[18-20,39]两种。

最著名的基于语义网络模型的相似性度量方法是由 Rada 等^[38]提出的,其基于路径距离的思想。概念表示为树结构的语义网络中的结点,由 Is-a 边连接,如概念 C 和 D 的相似性被计算为连接 C 和 D 的最短路径的长度^[35]。该方法的不足是:1)不能处理比 Is-a 更复杂的关系;2)对预定义的层次网络非常敏感,因而仅能够对具有相同祖先的概念进行粗略的相似性度量。Ontanon 等^[40]提出一种基于细化格的相似性度量方法。首先根据两个事例的索引概念的反合 AU(Anti-Unification,即最具体的泛化)来计算它们的共享信息,然后将相似性定义为它们的共享信息与它们的信息总和的比值。Sanchez-Ruiz 等^[41]将细化算子和细化图用于描述逻辑 ϵL 概念的相似性度量,提出基于个体相似性度量方法。该方法先计算两个个体的最具体概念 MSC,然后求两个 MSC 概念的最小公共包含概念 LCS,最后求取细化图中它们之间的距离。本质上,Recio-Garíá^[25]、Ontanon^[40]和 Sanchez-Ruiz^[41]的事例相似性度量方法都是基于语义网络模型的。该类方法存在的问题是:概念的反合 AU 和个体的最具体概念 MSC 的存在性

不能保证,或计算困难;通常只能给出它们的近似值,因此度量结果也不够准确。

基于描述逻辑范式的度量方法通过比较两个事例索引概念的语法描述来计算它们的相似性,并且是迭代定义的。计算过程需要利用本体形式的领域知识,而本体正是由描述逻辑来刻画的。对于两个待比较的事例索引概念,先给出它们的范式。范式是索引概念中基本概念合取式、值约束合取式或存在约束合取式等的析取形式。然后计算范式中各析取式中的合取式的相似性(即局部相似性),最后对各个部分的相似性结果加权求和,得出最终相似性(即全局相似性)。d'Amato^[20,42]给出描述逻辑 ALC 范式的定义,利用重叠函数定义了两个待比较概念的差异性。相似性可以通过差异性得到。在此基础上,d'Amato 等^[39]给出描述逻辑 ALN 范式的定义,提出了 ALCN 概念的相似性度量方法。Janowicz 等^[18]给出描述逻辑 ALCNR 范式的定义,提出了 ALCNR 概念的相似性度量方法。在此基础上,Janowicz 等^[19]给出描述逻辑 SHI 概念的相似性度量方法。

基于描述逻辑范式的度量方法全面地度量了事例索引概念的每个成分的相似性,并考虑了各个成分的重要程度,因此得出的相似性较为客观与准确。但其不足是:对于语义等价但语法形式不同的概念可能会给出不同的相似性结果,而且存在计算量大的问题。为了克服基于外延和基于内涵的度量方法的缺点,d'Amato 等^[43]提出了一种利用概念外延的思想,但既不属于基于外延又不属于基于内涵的语义相似性度量方法。该方法以两个概念的外延的实例数量相对于它们的公共超概念的外延的实例数量的变化作为相似性度量的依据。公共超概念通过概念的 GCS(Good Common Subsumer)^[44]来确定。该度量方法满足了等价可靠性、单调性准则,并且利用了概念的隐含语义,目前来说是一种更优良的数值相似性度量方法。

4.2.2 符号型度量方法

Salotti^[14]提出由于事例被表示成形式化的索引概念形式,在事例检索时可以定义事例间的符号形式的相似性和差异性。每个事例通过使用两个准则:相似性和差异性与目标事例进行比较。两个事例间的相似性由包含两个事例索引概念的最特殊概念 MSC 来刻画。两个事例间的差异性也由一个概念来刻画,这个概念表示了属于一个事例但不属于另一个事例的特征的集合。该方法的特点是显式地使用概念来表示两个事例间的相似性与差异性,以此可以得到一个事例的偏序序列,进而选出最相似的事例。不足之处是,该方法仅适用于具有最小公共包含概念 LCS、表达能力较弱的描述逻辑。对于强表达能力描述逻辑如 ALC,概念 C 和 D 的 LCS 是 CLD 或其等价概念,因此 LCS 不能表达有关概念 C 和 D 的相似特征的任何信息。Armengol 等^[45-47]提出了使用事例索引概念的合反 AU 的符号相似性度量方法,并将其应用于 CBR 事例分类任务。该方法使用符号来描述目标事例与不同类事例之间的共享特征。与数值型度量方法相比,符号型度量方法的优点是:能够用符号形式清晰地给出目标事例与检索事例间的相似性。通过这些符号相似性描述,用户可以了解事例的分类过程以及依据。其不足之处是不直观,不能直接给出多个事例间的相对相似程度。

借助于描述逻辑的推理能力,描述逻辑可以实现 CBR 事

例检索的正确性(确保检索事例满足搜索准则)、完备性(确保从事例库检索出目标事例的解)以及完成检索算法复杂性(检索算法是高效的,即多项式时间复杂性)等形式化证明。

5 基于描述逻辑的事例修正与事例库维护

在事例检索阶段得到的最相似事例可能不能直接用于解决目标问题或者在重用这个事例时产生矛盾,这时需要对其进行修正。目前,对基于描述逻辑的事例修正的研究已经取得了一定的进展。Gomez-Albarran 等^[16]提出了一个基于替换方法的通用修正方案。该方案可以识别出相似事例的解中应该被替换的成分,并指导搜索向最适合的替换方向进行。为了指出可以进行哪些替换,在修正过程中通过角色链来匹配源事例和目标事例的成分。即先用启发式搜索来查找被替换者;当启发式搜索没有找到被替换者时,再用搜索关系路径和检索个体的方法。该方案的缺点是:不能改变被修正的解的结构。这是因为显式地约束事例描述与其解之间的依赖关系限制了可能进行的修正操作。

Napoli 等^[48]将概念分类操作作用于基于框架表示的事例的修正过程,这也适用于描述逻辑。修正过程使用最小公共包含概念 LCS,充分利用分类层次来泛化相似事例的被替换成分。d'Aquin 等^[29]指出在事例修正阶段引入本体形式的领域知识为实现跨不同领域的事例修正重用的语义推理提供知识源。在多本体环境 CBR 的事例修正中,Recio-Garía 等^[25]给出了一个基于本体模型和修正规则的修正过程。对于给定的一类源事例,一个规则指定了相应的替换。

Cojan 等^[21,22]给出了描述逻辑 ALC 下的一种事例修正算法。该修正算法的原则是首先假定需要被修正的事例能够解决当前的目标事例。这可能会与目标事例的描述或领域知识发生冲突或逻辑矛盾,修正过程就是要修复这些冲突和矛盾。该算法使用经典 Tableau 方法的扩展,基于领域知识来实现修正操作。具体方法是:不断从源事例中暂时地删除断言知识,直到一致性恢复。Cojan 的进一步研究方向是将基于规则的修正算法与该算法结合起来,这样可以在需要时发现被替换者而不是删除导致冲突和矛盾的断言。

从以上看出,事例修正没有一种相对通用的方法。这使事例修正问题成了 CBR 研究中的一个难点。实际上,实现事例修正的源泉来自于具体的应用领域。基于描述逻辑的事例修正应更多地依赖于领域知识-修正知识来解决。d'Aquin 等^[28,29]研究了分散式 CBR 的多本体环境中的事例修正方法。

在目标问题得到解决后,为了解决以后遇到的类似问题,目标问题的解或解决方法应该被保存到事例库中。在基于描述逻辑的 CBR 中使用描述逻辑的推理机制,事例保存比较容易实现。Gomez-Albarran 等^[16]使用概念分类操作将事例索引保存到层次体系中,事例的索引概念被自动定位到正确的位置。但当事例索引的层次很深时,会带来事例检索效率的降低,因此不允许很深的索引层次。在利用分类操作搜索知识时,可以通过一个启发式搜索(已成功发现待修正解的被替换成分)并对它加上权重,来学习事例修正所用到的修正知识。

在基于描述逻辑的 CBR 的事例库中,事例库(知识库)的一致性检查很重要,特别是结构复杂的事例和规模比较大的

事例库。基于本体的 CBR 事例保存与事例库维护在文献 [25,28]中已被研究过。使用本体学习、概念聚类等方法可以更好地保存有用事例,控制事例库的规模,保证事例检索效率,从而稳定整个系统的性能。

6 基于描述逻辑的 CBR 目前存在的问题及未来的研究方向

CBR 是一种方法而非一种技术^[49],这种界定使 CBR 成为一个开放的系统。不断地吸收各种新的技术和方法更有利于 CBR 本身的发展。描述逻辑以其良好的语义和推理能力,成为 CBR 的理论基础之一。经过二十几年的努力,基于描述逻辑的 CBR 无论在基本理论、关键实现技术还是具体应用上都取得了长足的进展。但总体看来,目前还存在以下问题:

(1)目前用于 CBR 的描述逻辑未能引入一些实际应用领域中常见和重要的约束或语义(如数量约束、有型域、时序语义、模糊语义等),因而不能更准确地表示事例。这降低了事例检索的准确性,也影响了 CBR 系统的效率。

(2)目前大部分事例检索算法的计算量较大、效率不高,尤其是强表达能力的描述逻辑,如 SHOQ(D)描述的事例的相似性度量问题。

(3)目前的事例修正算法还不够完善,特别是对复杂事例的修正。

因此,未来基于描述逻辑的 CBR 研究的主要方向如下:

(1)根据应用领域的需要,将具有数量、有型域、时序或模糊等约束或语义的描述逻辑引入到 CBR 中,进一步提高事例描述的准确性和事例检索的可靠性与完备性。

(2)研究更高效的事例检索算法,复杂事例的相似性度量方法是重点之一。基于外延和内涵相结合的方法能更贴切地度量事例的相似性,这将是事例相似性度量的主要研究方向。结合 OBDD 符号技术^[50]来降低事例检索算法的复杂度是接下来需要考虑的问题。

(3)研究结合领域知识的事例修正算法。CBR 的关键技术,如事例表示、事例检索和事例修正等应尽可能地在该领域内寻求实现方案,尤其事例修正更应依赖于领域知识来解决。采用基于描述逻辑的本体表示领域知识,使其与基于描述逻辑的事例描述具有统一的形式,能提高事例修正的准确性与效率。修正规则、修正知识的使用也可以进一步提高事例修正的效率。因此引入特定的领域知识,结合规则和修正知识库将是实现事例修正的更优的方案。

(4)由于将来领域知识和系统资源有趋于分散的特点,分散(分布)式 KI-CBR 也将是未来的 CBR 研究方向之一。

参 考 文 献

- [1] Watson I. Applying case-based reasoning: techniques for enterprise systems[M]. Morgan Kaufmann Publishers Inc., 1998
- [2] Aamodt A, Plaza E. Case-based reasoning: Foundational issues, methodological variations, and system approaches[J]. AI Communications, 1994, 7(1): 39-59
- [3] Schank R C, Abelson R R. Goals and Understanding[M]. Erlbaum, Eksevier Science, 1977
- [4] Schank R C. Dynamic Memory: A Theory of Reminding and Learning in Computers and People[M]. Cambridge University Press, 1983
- [5] Kolodner J. Maintaining Organization in a Dynamic Long-term Memory[J]. Cognitive Science, 1983, 7(4): 243-280
- [6] Baader F, Calvanese D, McGuinness D. The Description Logic Handbook: Theory, Implementation and Applications[M]. Cambridge University Press, 2003
- [7] Katsuno H, Mendelzon A O. Propositional knowledge base revision and minimal change[J]. Artificial Intelligence, 1991, 52(3): 263-294
- [8] Richter M M. On the notion of similarity in case-based reasoning [J]. Mathematical and Statistical Methods in Artificial Intelligence, 1995: 171-184
- [9] Beck H W. A Conceptual Clustering Algorithm for Semantic Data Models[C]//International workshop on terminological logics, 1991. Schloss Dagstuhl, Germany, 1991: 17-22
- [10] Ashley K D, Alevin V. A Logical Representation for Relevance Criteria[C]// Althoff K-D, Weft S, Richter M M, et al., eds. Topics in Case-Based Reasoning. Springer Verlag, 1994: 338-352
- [11] Kolodner J. An Application of Terminological Logics to Case-based Reasoning[C]//Proc. KR94. Bonn, Germany, 1994
- [12] Kolodner J. Planning From Second Principles[J]. Artificial Intelligence, 1996, 87(1/2): 145-186
- [13] Kamp G. Using Description Logics for Knowledge Intensive Case-based Reasoning[C]//Lecture Notes in Computer Science 1168. Springer, 1996: 204-218
- [14] Salotti S, Ventos V. Study and Formalization of a Case-based Reasoning System Using a Description Logic [C] // Lecture Notes in Computers Science 1488. Springer, 1998: 286-297
- [15] Coupey P, Fouquere C, Salotti S. Formalizing Partial Matching and Similarity in Case-based Reasoning with a Description Logic [J]. Applied Artificial Intelligence, 1998, 12(1): 71-112
- [16] Gomez-Albarran M, Gonzalez-Calero P, et al. Modelling the CBR Life Cycle Using Description Logics [C] // Lecture Notes in Computers Science 1650. 1999: 147-161
- [17] Lenz M, Bartsch-Sporl B, et al. Case-based Reasoning Technology[C]//From Foundations to Applications, LNAI 1400. Springer Verlag, 1998
- [18] Janowicz K. Sim-DL: Towards a Semantic Similarity Theory for the Description Logic ALCNR in Geographic Information Retrieval[C]//Lecture Notes in Computers Science 4278. Springer, 2006: 1681-1692
- [19] Janowicz K, Wilkes M. SIM-DLA: A Novel Semantic Similarity Measure for Description Logics Reducing Interconcept to Interinstance Similarity[C]//Lecture Notes in Computers Science 5554. Springer, 2009: 353-367
- [20] d'Amato C, Fanizzi N, Esposito F. A Dissimilarity Measure for the ALC Description Logic[C]//Proceedings of the 2nd Italian Semantic Web Workshop. 2005
- [21] Cojan J, Lieber J. An Algorithm for Adapting Cases Represented in an Expressive Description Logic[C]//Lecture Notes in Computers Science 6176. 2010: 51-65
- [22] Cojan J, Lieber J. An Algorithm for Adapting Cases Represented in ALC[C]//Proceedings of the 22nd International Joint Conference on Artificial Intelligence. Barcelona, Spain, 2011: 2582-2589

- ment problem in program understanding[C]// Proceedings of the 15th International Conference on Software Engineering. IEEE Computer Society Press, 1993; 482-498
- [2] Dit B, Reville M, Gethers M, et al. Feature Location in Source Code: A Taxonomy and Survey[J]. *Journal of Software: Evolution and Process*, 2013, 25(1): 53-95
- [3] Trifu M. Using Dataflow Information for Concern Identification in Object-Oriented Software Systems[C]// Proceedings of European Conference on Software Maintenance and Reengineering (CSMR' 08). 2008; 193-202
- [4] Robillard M P. Topology Analysis of Software Dependencies [J]. *ACM Transactions on Software Engineering and Methodology*, 2008, 17(4): 1-36
- [5] Wong W E, Gokhale S S, Horgan J R, et al. Locating program features using execution slices[C]// Proceedings of IEEE Symposium on Application-Specific Systems and Software Engineering and Technology (ASSET'99), March 1999; 194 -203
- [6] Eisenberg A D, De Volder K. Dynamic Feature Traces: Finding Features in Unfamiliar Code[C]// Proceedings of 21st IEEE International Conference on Software Maintenance (ICSM'05). Budapest, Hungary, September 2005; 337-346
- [7] Zhao W, Zhang L, Liu Y, et al. SNIAFL: Towards a Static Non-interactive Approach to Feature Location[J]. *ACM Transactions on Software Engineering and Methodologies (TOSEM)*. 2006, 15(2): 195-226

(上接第 6 页)

- [23] Aamodt A. Knowledge-Intensive Case-Based Reasoning and Sustained Learning[C]// ECAI. 1990, 90
- [24] Aamodt A. Knowledge-intensive case-based reasoning in creek [M]// *Advances in Case-Based Reasoning*. Springer Berlin Heidelberg, 2004; 1-15
- [25] Recio-Gara J A, Dıaz-Agudo B. Ontology based CBR with jCOLIBRI[M]// *Applications and Innovations in Intelligent Systems XIV*. Springer London, 2007; 149-162
- [26] Juarez J M, Salort J, Palma J, et al. Case Representation Ontology for Case Retrieval Systems in Medical Domains[C]// *Proceeding of the IASTED International Conference on Artificial Intelligence and Applications (AIA2007)*, 2007; 168-173
- [27] Plaza E, McGinty L. Distributed case-based reasoning[J]. *The Knowledge engineering review*, 2005, 20(3): 261-265
- [28] d' Aquin M, Lieber J, Napoli A. Decentralized case-based reasoning for the Semantic Web[M]// *The Semantic Web-ISWC 2005*. Springer Berlin Heidelberg, 2005; 142-155
- [29] d' Aquin M, Lieber J, Napoli A. Decentralized case-based reasoning and Semantic Web technologies applied to decision support in oncology[J]. *The Knowledge Engineering Review*, 2013, 28(4): 425-449
- [30] Bergmann R, Kolodner J, Plaza E. Representation in case-based reasoning[J]. *The Knowledge Engineering Review*, 2005, 20(3): 209-213
- [31] de Mantaras R L. *Case-based Reasoning*[M]. Springer-Verlag, Heidelberg, 2001
- [32] Cao M J, Shang F H, et al. Research on Scalable Case Representation and its Retrieval Based on Description Logic[C]// *Second International Symposium on Knowledge Acquisition and Modeling*. 2009; 316-318
- [33] Lieber J, Napoli A. Using Classification in Case-Based Planning' [C]// *ECAI PITMAN*, 1996; 132-136
- [34] Hunter J, Drennan J, Little S. Realizing the Hydrogen Economy Through Semantic Web Technologies[J]. *Intelligent Systems, IEEE*, 2004, 19(1): 40-47
- [35] Tversky A. Features of similarity [J]. *Psychological Review*, 1977, 84; 327-352
- [36] Borgida A, Walsh T, Hirsh H. Towards Measuring Similarity in Description Logics[C]// *Proceedings of the 2005 International Workshop on Description Logics*. Edinburgh, Scotland, UK, 2005
- [37] d' Amato C, Fanizzi N, Esposito F. A Semantic Similarity Measure for Expressive Description Logics[C]// *Proceedings of Convegno Italiano di Logica Computazionale*. Rome, Italy, 2005
- [38] Rada R, Mili H, Bicknell E, et al. Development and Application of a Metric on Semantic Nets[J]. *IEEE Transactions on Systems, Man and Cybernetics*, 1989, 19(1): 17-30
- [39] d' Amato C, Fanizzi N, Esposito F. A Semantic Similarity Measure for ALN Description Logics[C]// *Proceedings of Convegno Italiano di Logica Computazionale*. Rome, Italy, 2005
- [40] Otonon S, Plaza E. On similarity Based on a Refinement Lattice [C]// *Lecture Notes in Computer Science*, 5650. Springer, 2009; 240-255
- [41] Sanchez-Ruiz-Granados A A, Ontanon S, Gonzalez-Calero P A, et al. Measuring Similarity in Description Logics Using Refinement Operators[C]// *Lecture Notes in Artificial Intelligence*, 6880. Springer, 2011; 289-303
- [42] d' Amato C, Fanizzi N, Esposito F. Analogical reasoning in description logics[M]// *Uncertainty reasoning for the Semantic Web I*. Springer Berlin Heidelberg, 2008; 330-347
- [43] d' Amato C, Staab S, Fanizzi N. On the Influence of Description Logics Ontologies on Conceptual Similarity[C]// *Lecture Notes in Computer Science*, 5268. Springer, 2008; 48-63
- [44] Baader F, Sertkaya B, Turhan A Y. Computing the Least Common Subsumer wrt a Background Terminology[J]. *Journal of Applied Logic*, 2007, 5(3): 392-420
- [45] Armengol E, Plaza E. Using Symbolic Descriptions to Explain Similarity on CBR [C]// *Proceedings for the 8th International Conference of the ACIA*. Alguer, Italy, 2005; 239-246
- [46] Armengol E, Plaza E. Symbolic Explanation of Similarities in Case-based Reasoning[J]. *Computers and Artificial Intelligence*, 2006, 25(2/3): 153-171
- [47] Plaza E, Armengol E, Ontaon S. The Explanatory Power of Symbolic Similarity in Case-based Reasoning[J]. *Artificial Intelligence Review*, 2005, 24(2): 145-161
- [48] Napoli A, Lieber J, Courien R. Classification-based Problem Solving in CBR[M]// *Smith I. & Faltings B., eds. Advances in CBR*, Springer-Verlag, 1996
- [49] Watson I. Case-based Reasoning is a Methodology, not a Technology[J]. *Knowledge-based Systems*, 1999, 12(5/6): 303-308
- [50] Torta G, Torasso P. The role of OBDDs in controlling the complexity of model based diagnosis[C]// *Fifteenth international workshop on principles of diagnosis (DX'04)*, 2004