



西方哲学研究丛书

信念—愿望—意图逻辑 及其应用研究

张晓君 著

Research on Belief-Desire-Intention Logic
and its Applications

Research on Belief-Desire-Intention Logic
and its Applications

中国社会科学出版社



西方哲学研究丛书

信念—愿望—意图逻辑 及其应用研究

张晓君 著

Research on Belief—Desire—Intention Logic
and its Applications

中国社会科学出版社

图书在版编目(CIP)数据

信念—愿望—意图逻辑及其应用研究 / 张晓君著. —北京: 中国社会科学出版社, 2017. 9

ISBN 978-7-5161-9398-3

I. ①信… II. ①张… III. ①逻辑学 IV. ①B81

中国版本图书馆 CIP 数据核字(2016)第 288146 号

出 版 人 赵剑英
责任编辑 冯春风
责任校对 张爱华
责任印制 张雪娇

出 版 中国社会科学出版社
社 址 北京鼓楼西大街甲 158 号
邮 编 100720
网 址 <http://www.csspw.cn>
发 行 部 010-84083685
门 市 部 010-84029450
经 销 新华书店及其他书店

印 刷 北京君升印刷有限公司
装 订 廊坊市广阳区广增装订厂
版 次 2017 年 9 月第 1 版
印 次 2017 年 9 月第 1 次印刷

开 本 710 × 1000 1/16
印 张 15
插 页 2
字 数 213 千字
定 价 68.00 元

凡购买中国社会科学出版社图书,如有质量问题请与本社营销中心联系调换
电话:010-84083683
版权所有 侵权必究

本专著是以下课题的阶段成果：

2013 年教育部人文社会科学研究规划基金项目“基于 BDI（信念、愿望和意图）逻辑的慎思 Agent 的行为决策研究（编号：13YJA72040001）”

序 言

周昌乐

(福建省仿脑智能系统重点实验室主任、
厦门大学教授、博士生导师)

看到张晓君新著《信念—愿望—意图逻辑及其应用研究》即将出版,感到非常高兴。作为她的博士后指导老师,了解该书是在其博士后出站报告的基础上修改完善而成的。当年她的这部博士后出站报告被评为厦门大学优秀博士后出站报告,印象非常深刻。可以说,这是国内第一部关于信念—愿望—意图逻辑研究的专著。

作为认知型智能 Agent (主体),其典型的行为模型就是 BDI (信念—愿望—意图) 逻辑框架。BDI 主体模型在一些最为重要的多主体系统也得到了广泛应用,是影响最为广泛的 Agent 技术之一。张晓君的这部专著,以 Agent 不确定性行为决策和推理的建模为主要研究内容。“行为”或“行动”是哲学、心理学、历史、法律、经济、军事、计算机科学、系统工程、人工智能等诸多学科研究的核心概念,这是因为行动在这些领域中无处不在。比如,就机器而言,一个程序的执行和调用本质上就是一个行动,并据此改变机器的内部世界,然后通过与外部世界的接口来改变外部世界。对于具有自我意识的机器人而言:行动就是通过感知信息控制执行过程的算法。当然,对机器人的拟人行为的研究是人工智能的核心、关键和高难度课题,其行为建模涉及的领域非常广泛。单从这个角度上看,张晓君的这部专著不仅能够促进智能 Agent 不确定性行为的表示和推理,而且能够

为先进分布式仿真、人工智能和智能机器人等领域的复杂行为建模,提供形式化支持,因此具有重要的理论价值和实践意义。

从这部专著的具体指导思想和内容上看,其研究的深度和难度在国内这一领域都是领先的,包括了诸多创新性成果,主要体现在以下几个方面:(1)提出了等级 BDI 逻辑 $g\text{-BDI}_{\text{PDL+LL}}$,对信念、愿望和意图的模型进行了公理化,并说明了它们对 Agent 行为的影响;(2)提出了情感等级 BDI 逻辑,并对其进行了公理化,清晰地表示了 Agent 的心理状态和情感状态的不确定性,说明了它们对 Agent 行为的影响;(3)把无穷值 Łukasiewicz 逻辑的真值取值范围从 $[0, 1]$ 扩展到 $[-1, 1]$,并提出了知识情感等级 BDI 逻辑,清晰地描述了 Agent 的知识状态、心智状态和情感状态的不确定性。我相信,所有这些创新性成果对国内学术界的信念、愿望和意图逻辑及其应用研究将产生较为重要的影响,并有助于促进国内人工智能逻辑的进一步深入发展。

总之,从总体上看,这部专著是从 Agent 的信念、愿望、意图、甚至知识、情感等知识状态和情感状态出发,对 Agent 的行为决策和行为推理进行建模。整部专著立论新颖独到,参考资料翔实可靠,论述结构合理、条理清晰、逻辑严密,表述流畅,引用规范,是一部纯粹的学术专著。衷心期待这部专著的出版,能够引起学术界的关注,并为繁荣国内人工智能逻辑学研究做出贡献。

2015 年 7 月 28 日于厦门

前 言

做知识级分析的系统称为 Agent（主体，也译为“智能主体”、“智能体”、“真体”等，比如人、具有智能的高级动物和仿人机器人均可视为智能主体）。Agent 强调理性作用，是描述人类智能、动物智能和机器智能的统一模型，这些智能往往是构成社会智能的一部分。Agent 是一种通过传感器感知其环境，再通过执行器作用于该环境的智能实体，实则是从感知序列到实体动作的映射。这一定义使得我们可以通过函数和数值对行动或行为进行建模。行为建模涉及的领域非常广泛，比如系统分析与设计、机械设计、动画设计、网络安全、军事仿真、仿人机器人等。因此，对行为（尤其是不确定性行为）进行精细的逻辑刻画具有重要的理论价值和实践意义。

Agent 行为理论将 Agent 定义为有意识系统，即 Agent 的行为可通过归因于信念、期望和理性的方法去预言。经典的一阶逻辑不能用来表示涉及意识的观念，否则会引起语法和语义问题。引入模态逻辑和元语言可解决语法问题；借助可能世界语义和解释的符号结构则可解决语义问题。BDI（Belief-Desire-Intension，信念—愿望—意图）主体是指其行为受到信念、愿望和意图支配的主体。Agent 典型的行为模型就是 BDI 逻辑框架，它着重研究信念、愿望、意图与 Agent 行为的关系及其形式化描述。

基于 BDI 主体的行为建模方法主要有两种：一是从逻辑学的视角来研究，严格遵循 BDI 的形式化语义，通过定义不同的逻辑算子来建立主体的个体模型；二是从人工智能和系统工程的可实现性的视

角来研究,把主体看作面向对象技术中的对象的扩展。本专著主要从逻辑学的视角来论述。

从逻辑学的视角来看,BDI 模型的研究主要有三种方法:(1)以分支时态逻辑为基础的方法;(2)以线性时态逻辑为基础的方法;(3)以命题动态逻辑为基础的方法。其中,影响最为深远的是 Rao 和 Georgeff (1991)提出的以分支时态逻辑为基础的 BDI 主体模型。近年来,不少学者对这一模型进行了改进和扩展,其中较为重要的有:(1)情感(emotional)BDI 主体模型;(2)等级(graded)BDI 主体模型;(3)情感等级 BDI 主体模型;(4)知识情感等级 BDI 主体模型。

由于 BDI 逻辑系统易于在计算机中得以实现,所以,备受计算机科学、人工智能、系统工程等领域相关专家的青睐。BDI 主体模型已经在这些领域得到了广泛的应用。例如,基于 BDI 的主体模型已经在机场空中交通管制(Busetta et al., 2003)、计算机生成虚拟兵力(陈中祥, 2004)、多主体系统的机器学习(Hernández et al., 2004)、火星机器人岩石样本的选择(Bordini and Hübner, 2006)、铁路货运装车日计划审批(刘刚等, 2009)等多个领域获得了应用。目前,BDI 主体模型的研究在这些领域中正如火如荼地开展着,例如, Lincoln 与 Veres (2012)已经给出了复杂机器 BDI 主体的自然语言程序设计。近年来,计算机科学界和人工智能界等领域的不少学者已经使用 Java、Jason、JaCa 等程序设计语言对 BDI 主体模型进行了计算机实现。笔者之所以从事信念—愿望—意图逻辑研究,正是因为该逻辑具有重要的理论价值和应用价值。

本专著是在笔者的博士后出站报告《基于 BDI (Belief-Desire-Intention) 逻辑及其扩展的 Agent 行为表征研究》的基础上修改完善而成的。其内容主要包括国内外相关研究现状、基于 BDI 逻辑的 Agent 行为表征研究、基于扩展 BDI 逻辑的 Agent 行为表征研究、基于扩展 BDI 逻辑的 Agent 行为规范研究。笔者的创新部分主要体现在第三章和附录 1—附录 3 中。

笔者在厦门大学做博士后期间，认真学习了智能科学与技术系的周昌乐教授为博士生开设的《人工智能前沿》课程；全程学习了史晓东教授为智能科学与技术系的博士生开设的《计算语言学》课程；仔细聆听了周昌乐教授为全校学生开设的“沟通文理，启迪智慧”的系列讲座；认真听取了厦门大学信息科学学院主办的各类中外专家讲座、国际国内会议 10 余次，重点关注逻辑学与计算机科学交叉研究的契合点，寻找逻辑学研究实用化的突破口；积极参与了福建省仿脑系统重点实验室（厦门大学）每两周一次的报告，并为该实验室的成员作报告，重点关注人工智能科学对逻辑学的重点需求领域，努力探索和思考二者交叉研究的结合点和着力点，为专著写作寻找思路和灵感。但即便如此，“书到用时方恨少”，再加之“光阴似箭，日月如梭”，纰漏和瑕疵在所难免，敬请各位专家和学者不吝赐教。

张晓君

2015 年仲夏于成都

摘 要

“行为”是哲学、经济、军事、计算机科学、系统工程、人工智能等诸多学科研究的核心概念，这是因为它们在这些领域中无处不在。行动实际上就是智能主体（Agent）通过感知信息控制执行过程的算法。智能主体典型的行为模型就是 BDI（信念—愿望—意图）逻辑框架。BDI 主体模型在一些最为重要的多主体系统中得到了广泛应用，是影响最为广泛的主体技术之一。

本专著以主体不确定性行为决策和推理为主要研究内容。本研究主要以 Pereira 等人（2008）、Casali 等（2005, 2011a, 2011b）、Rao 和 Georgeff（1991, 1995, 1998）、Criado 等（2010）、Meyer 等（2012）、Galind 和 Saffiotti（2013）、Puică 和 Florea（2013）等相关文献为基础，借鉴其研究思路和研究方法，并对其研究成果进行了综合、拓展、延伸和创新。本研究能够促进智能主体不确定性行为的表示和推理，并能够为先进分布式仿真、人工智能和系统工程等领域提供形式支持，具有重要的理论价值和实践意义。本专著安排如下：

第一章是 BDI 逻辑及其扩展的研究与其应用概况介绍；第二章是基于 BDI 逻辑的主体行为表征研究；第三章是基于扩展 BDI 逻辑的主体行为表征研究；第四章是基于扩展的 BDI 逻辑的主体行为规范研究；每章都对其相应研究成果进行了实例分析。

第一章和第二章是本专著的思想基础和技术基础；第三、四章是在前两章基础上的拓展研究。第三章是本专著研究的重点内容和创新内容。

第三章内容属于国际研究前沿，是有别于国内外学者工作的特色研究，笔者的创新内容主要体现在这一章中。这些创新主要体现在以下几个方面：

(1) 把无穷值的 Łukasiewicz 逻辑和命题动态逻辑进行融合后，在对等级 BDI 主体模型进行形式化的基础上，笔者提出了等级 BDI 逻辑 $g\text{-BDI}_{\text{PDL}+\text{LL}}$ ，对信念、愿望和意图的模型进行了公理化，并说明了它们对主体行为的影响。

(2) 为了通过信念度、愿望度、意图度、害怕度、焦虑度和自信度对不确定性行为进行表示和推理，笔者提出了情感等级 BDI 逻辑，并对其进行了公理化，清晰地表示了主体的心理状态和情感状态的不确定性，说明了它们对主体行为的影响。

(3) 为了对知识、信念、愿望和意图以及害怕、焦虑和自信对主体决策行为影响的强烈程度进行形式化，笔者把无穷值的 Łukasiewicz 逻辑的真值取值范围从 $[0, 1]$ 扩展到 $[-1, 1]$ ，并提出了知识情感等级 BDI 逻辑，清晰地描述了主体的知识状态、心智状态和情感状态的不确定性。笔者提出的这些模型可以轻易地向其他类型的主体模型进行推广。

关键词：BDI（信念—愿望—意图）逻辑；知识状态；心智状态；情感状态；主体行为；不确定性推理

Abstract

Behavior is a core concept in philosophy, economy, military, computer science, systems engineering, artificial intelligence, and among many other fields, due to the ubiquitousness of behavior in these fields. In fact actions are algorithms that agents control executive procedures by perceptive information. The BDI (Belief-Desire-Intention) logical framework is a typical behavioral model of agents. BDI agent architecture has been applied in some of the most significant multi-agent systems developed up to now, and has been one of the most influential theories with respect to agent technology.

This monograph is mainly concerned with decision-making and reasoning of agents' uncertain behavior. It draws chiefly on the ideas and methods in Pereira et al. (2008), Casali et al. (2005, 2011a, 2011b), Rao & Georgeff (1991, 1995, 1998), Criado et al. (2010), Meyer et al. (2012), Galind & Saffiotti (2013), and Puică & Florea (2013), and among any others. It attempts not only to synthesize and expand on past achievements but also to make novel and original proposals. It is hoped that the present study will make contributions to represent and reason agents' uncertain behavior as well as providing a formal support for advanced distributed simulation, artificial intelligence, systems engineering, and among many other fields. The organization of the monograph is as follows:

Chapter One profiles the research on BDI logic and its extensions and

their application. Chapter Two presents characterization of agents' behavior based on BDI logic. And characterization of agents' behavior based on extended BDI logics is illustrated in Chapter Three. Finally, Chapter Four discusses behavioral norms based on extended BDI logics. Some applications on the basis of these studies are given in the corresponding chapters.

The first and second chapters lay the conceptual and technical bases of the whole monograph. The third and fourth chapters are extensions on the basis of the first and second chapters. The third chapter is the focus of the whole monograph, with carrying the biggest weighting in originality.

The whole Chapter 3 reports original findings which are believed to be at the very forefront of the research in the logic circle worldwide. The major achievements are summarized as follows:

(1) On the basis of blending the infinite-valued Łukasiewicz Logic and Propositional Dynamic Logic to formulize the graded BDI agent model, the author proposes a graded BDI logic ($g\text{-BDI}_{\text{PDL} + \text{LL}}$ for short) in this paper, and looks for a possible axiomatic modeling of beliefs, desires and intentions, and then shows how they influence agents' behavior.

(2) In order to represent the uncertainty behavior by belief, desire, intention, fear, anxiety and self-confidence degree, the corresponding axioms are added to the Łukasiewicz logic. The emotional graded BDI logic proposed by the author explicitly represents the uncertainty of mental attitudes and emotional states. The paper is to axiomatize the model of beliefs, desires, intentions, fear, anxiety and self-confidence, and to show how they influence agents' behavior.

(3) In order to formalize the degree of knowledge, belief, desire, intention, fear, anxiety and self-confidence which influence on agents' decision-making behavior, the author extends the truth value of infinite-valued Łukasiewicz Logic from $[0, 1]$ to $[-1, 1]$, and then proposes the knowledge of emotional graded BDI logic, with explicitly representing the

uncertainty of knowledge states, mental attitudes and emotional states. These models proposed by the author in this paper are general enough to specify other different types of agents.

Key words: Belief–Desire–Intention logic; knowledge states; mental attitudes; emotional states; agents' behavior; uncertain reasoning

目 录

第一章 绪 论	(1)
第一节 研究背景及研究意义	(1)
第二节 国内外研究现状	(5)
2.1 BDI 逻辑的产生与发展	(5)
2.2 扩展的 BDI 逻辑的产生与发展	(7)
第三节 研究目标、主要内容、基本思路和方法、重点 难点、主要观点和创新	(11)
3.1 研究目标	(11)
3.2 主要内容	(12)
3.3 基本思路与方法	(13)
3.4 重点与难点	(15)
3.5 主要观点	(16)
3.6 主要创新	(22)
第四节 信息时代呼唤更多的连续值柔性高等逻辑的诞生	(23)
本章小结	(25)
第二章 基于 BDI (信念—愿望—意图) 逻辑的 Agent 行为表征研究	(27)
第一节 Bratman 的信念—愿望—意图理论	(28)
第二节 Cohen 和 Levesque 关于意图的方法	(30)
2.1 行动与时间	(30)
2.2 信念和偏好	(33)
2.3 行动、时间、信念和偏好逻辑	(34)

2.4 行动、时间、信念和偏好逻辑的扩展研究	(35)
第三节 Rao 和 Georgeff 的 BDI 逻辑	(38)
3.1 非形式语义	(38)
3.2 形式理论	(41)
3.3 作为变化公理的承诺	(51)
3.4 逻辑性质	(55)
第四节 KARO 逻辑	(57)
4.1 关于理性主体的 KARO 逻辑	(58)
4.2 关于情感主体的 KARO 逻辑	(64)
第五节 stit 逻辑中的 BDI-模态词	(67)
5.1 瞬间 stit 逻辑中的 BDI 模态词	(67)
5.2 在 XSTIT 逻辑中的 BDI 模态词: 动态视角	(74)
本章小结	(78)
第三章 基于扩展 BDI (信念—愿望—意图) 逻辑的 Agent	
行为表征研究	(81)
第一节 基于等级 BDI 逻辑的行为表征研究	(81)
1.1 引言	(81)
1.2 等级 BDI 逻辑	(83)
1.3 等级 BDI 逻辑对主体复合行动的形式刻画	(90)
1.4 实例分析: 等级 BDI 逻辑对主体买房行为的 形式刻画	(92)
1.5 结论和未来的工作	(95)
第二节 基于情感等级 BDI 逻辑的行为表征研究	(96)
2.1 引言	(96)
2.2 情感等级 BDI 逻辑	(97)
2.3 情感等级 BDI 逻辑对主体模型不同背景之间的 相互关系的形式刻画	(103)
2.4 实例分析: 情感等级 BDI 逻辑对主体买房行为的 形式刻画	(105)

2.5 结论和未来的工作	(109)
第三节 基于知识情感等级 BDI 逻辑的行为表征研究	(110)
3.1 引言	(110)
3.2 知识情感等级 BDI 逻辑	(111)
3.3 知识情感等级 BDI 逻辑对主体模型的不同背景之间的 相互关系的形式刻画	(116)
3.4 实例分析：知识情感等级 BDI 逻辑对军事决策行为的 形式刻画	(118)
3.5 结论和未来的工作	(123)
本章小结	(124)
第四章 基于扩展 BDI (信念—愿望—意图) 逻辑的 Agent 行为规范研究	(127)
第一节 引言	(128)
第二节 相关基础知识：多背景 BDI	(129)
2.1 BDI 背景	(129)
2.2 桥规则	(131)
第三节 规范模型方案	(132)
第四节 规范 BDI 结构	(134)
4.1 识别背景	(135)
4.2 规范背景	(135)
4.3 规范桥规则	(136)
第五节 实例研究的执行	(139)
第六节 思考与未来的工作	(141)
本章小结	(143)
结束语	(145)
参考文献	(147)
附录 1: Graded BDI Models for Agent Architectures Based on Łukasiewicz Logic and Propositional Dynamic Logic	Xiaojun Zhang et al. (160)