

智能体的理论研究

The Research on Agent Theory

毛新军 陈火旺 王怀民 齐治昌

(国防科技大学计算机系 长沙410073)

摘 要 Agent has become an important concept in AI/DAI and current mainstream computer disciplines. The aims of the research on agent theory are to make it clear what is agent, analyse, specify and verify the agent's properties. This paper discusses what is agent from the standpoint of theory research, analyses the formal tool used in the agent's theory research, reviews the current status and finally demonstrates the problems existed and the future research directions.

关键词 Agent, Possible world, Intentional stance

智能体在 AI/DAI 和当前主流计算机学科(如分布式计算)中正成为一个广泛使用、日益重要的关键性概念,由此而产生了一系列新的思想、方法、技术和语言用以规范,验证,设计和实现复杂的分布式软件系统,如基于智能体的软件工程(ABSE),面向智能体的程序设计(AOP),智能体通讯语言(ACL)等等^[1]。这些思想、方法、技术和语言的提出和应用极大地丰富了软件开发方法和技术,促进了复杂分布式软件系统的开发,推动了相关主流计算机学科的发展。今天,软件系统日益复杂和庞大,基于智能体的计算(Agent-based Computing)被认为是促进软件开发、克服软件危机的一个重要突破口,因而有关智能体理论和技术的研究和应用在学术界和工业界引起了人们的高度重视和极大兴趣。

智能体的理论研究可追溯到60年代,当时的研究侧重于讨论作为信息载体的智能体在描述信息和知识方面所具有的特性。直到80年代中后期,由于智能体技术的广泛应用以及在实际应用中面临的种种问题,智能体的理论研究才得到人们的重视,并于1993年首次召开了智能体形式化模型的国际会议,于1994年召开了第一届智能体理论、体系结构和语言的国际会议。智能体的理论研究试图解决三方面的问题:(1)什么是智能体?(2)智能体具有哪些特性?(3)如何采用形式化的方式描述和研究智能体的这些特性?它的研究旨在澄清智能体的概念,分析、

描述和验证智能体的有关特性,从而来指导智能体体系结构和 AOP 的设计和研发,促进复杂软件系统的开发。

本文旨在强调智能体理论研究的重要性和意义,讨论智能体的概念,分析智能体理论研究所需的形式化工具,系统地概述当前的研究状况和已有的成果,并就存在的问题和进一步研究阐述了我们的观点和看法。

1. 什么是智能体

同智能概念问题一样,智能体概念亦是目前引起人们广泛争论的问题之一。一方面,智能体是当前使用最为广泛的术语之一,但其含义通常随应用环境的变化而不同,要形成一个统一的、为大家所接受的概念是非常困难的;另一方面,智能体通常被认为具有某种智能特征,但对于什么是智能目前仍无一致的认识。在 AI 领域,一个能为大部分研究人员所接受的定义是将智能体看作是在某一环境中持续自主发挥作用、有生命周期的计算实体。尽管目前尚无非常确切的智能体的概念定义,但一种普遍的观点认为,作为智能体的软件或硬件系统一般具有以下特征:

• 自主性:智能体能在没有人类或其他智能体直接干涉和指导的情况下持续运行,并能控制其内部状态和动作。这是智能体区别于对象(object)的一个重要特征;

Oriented Analysis and Database Design, Information and Software Technology, 34(4)1992

[4] J. V. D. Bussche et al., Using SQL with Object-Orient-

ed Database, Information Systems, 18(7)1993

[5] 严开涛,何斌,基于 STEP 的质量管理信息系统,计算机工程与应用, 32(6)1996

• 社会性:智能体拥有其他智能体(可能是人类)的信息和知识,并能通过某种智能体通讯语言与其他智能体进行交互、协同和合作;

• 反应性:智能体能感知环境,并能对环境中的事件作出反应;

• 理性:智能体没有冲突的目标,其动作和行为总是基于其已有的目标,而且不会故意阻止其目标的实现;

• 忠诚性:智能体的通讯内容从不涉及为假的信息;

• 友好性:智能体总是尽可能地完成其他智能体的要求。

AI 领域的研究人员通常基于一种“意向姿态(intentional stance)”的观点来研究智能体的概念,分析和描述智能体的智能特征。所谓“意向姿态”观点是指将通常应用于人类自身的一些认知思想和概念(如信念、意愿、承诺)应用于智能体。对于计算机科学而言运用“意向姿态”的观点来研究智能体是极为有用的,因为它能帮助我们避开智能问题的争论,刻画和描述智能体的概念和特征,认识智能体的结构,分析、理解和预测智能体的行为以及智能体间的交互、通讯、协同和合作。例如,张三努力工作是因为他想(体现了一种意愿)得到晋升,李四穿上了棉袄是因为他认为(体现了一种信念)天气会转冷,张三请求(体现一种交互和合作)李四帮他学习英语是因为他想掌握英语。

根据各个认知概念的功能和特点,我们把智能体的认知状态分为三类:信息类别、动机类别和决策类别(见表1)。

表1

信息类别	信念、知识
动机类别	期望、意愿、承诺、选择等
决策类别	规划、know-how 等

目前人们仍不清楚上述认知状态是否足以认识智能体概念,哪些认识状态对于智能体而言是重要和必需的,哪些认知状态的组合对于智能体是比较恰当和合适的。文[5]认为智能体是信念(belief)-期望(desire)-意愿(intention)(BDI)三元认知结构。文[7,8]则认为信念、意愿和 know-how 是智能体三个重要的科学抽象。我们的观点是智能体由各个类别上至少一个认知状态来表示是比较恰当和合适的。此外,各个类别间、每个类别的各个认知状态间具有哪些关系仍是目前人们感兴趣的研究课题。已有的

研究表明信息类别和动机类别间、信息类别和决策类别间是相关联的,动机类别和决策类别间则是相互正交的^[6]。在信息类别和动机类别内部,许多认知状态是相互联系的^[6,7-9]。智能体理论研究旨在严密、形式化地讨论各个认知状态的概念,研究智能体的认知结构,分析智能体的性质和特征。

2. 智能体研究的形式化工具

经典的命题和一阶逻辑系统不适合于智能体的理论研究^[1],在 AI 领域,智能体的理论研究通常基于正规模态逻辑。正规模态逻辑的语义基础是 Kripke 可能世界语义,它的语义模型通常是指三元偶 $M = \langle U, R, I \rangle$ 。其中 U 为可能世界的集合, U 中的元素代表各个可能世界; R 为 U 上的二元关系,称为可能世界间的可达关系; I 是解释。智能体的认知状态由相应的模态算子来表示。通过引入相应的可到达关系 R ,我们可以基于语义模型定义模态算子的语义,获取认知状态的一些重要的性质。例如,用模态算子 Bef 表示智能体的信念,基于语义模型 M 我们通常将 Bef 的语义定义为:

$$M, s \models Bef \varphi \text{ iff } \forall s' \in U ((s, s') \in R \Rightarrow M, s' \models \varphi)$$

(1)

其中, φ 为一命题。直观地,智能体具有某种信念 φ 当且仅当在智能体认为所有与当前世界相关的可能世界 s' 中,命题 φ 为真。我们可对模型中的关系 R 作出进一步的约束,从而可以获得智能体的一些重要的特性。例如,限定 R 是连续的,则有公理 $D: Bef\varphi \rightarrow Bef\rightarrow\varphi$ 。公理 D 指出智能体的信念是一致的。限定 R 是自反的,则得到公理 $T: Bef\varphi \rightarrow \varphi$ 。公理 T 又称知识公理,它指出如果智能体具有某个信念 φ ,则 φ 为真。公理 T 是区分知识和信念的一个重要标志。因为公理 T 是智能体的知识所必须的,而智能体的信念通常不具备这种性质,所以公理 T 通常不作为智能体信念公理系统的组成部分。研究可能世界可达关系 R 所具有的性质(如连续,自反等)与相应公理系统间的对应关系称为正规模态逻辑的对称理论。

正规模态逻辑存在逻辑“无所不晓”(Logical Omniscience)问题。这一问题以及由此带来的其他一系列相关问题(如副作用问题、传递问题等)给智能体的理论研究带来了困难。考虑以下几个在正规模态逻辑系统中为真的命题。其中 φ 和 Ψ 为命题, X 为表示智能体某种认知状态的模态词。

$$(L01) \models X\varphi \wedge X(\varphi \rightarrow \Psi) \rightarrow X\Psi$$

$$(L02) \models \varphi \Rightarrow \models X\varphi$$

$$(L03) \models (\varphi \rightarrow \Psi) \Rightarrow \models (X\varphi \rightarrow X\Psi)$$

$$(L04) \models (\varphi \leftrightarrow \Psi) \Rightarrow \models X(\varphi \leftrightarrow \Psi)$$

$$(L05) \models (X\varphi \wedge X\Psi) \rightarrow X(\varphi \wedge \Psi)$$

$$(L06) \models X\varphi \rightarrow X(\varphi \vee \Psi)$$

$$(L07) \models \neg(X\varphi \wedge X\neg\varphi)$$

其中, (L01) 为 K 公理, (L02) 为必然规则。 (L01) 和 (L03) 构成了正规模态逻辑系统的副作用问题, 即逻辑系统关于智能体的认知状态封闭。 (L02) 构成了正规模态逻辑系统的传递问题。上述属性是对理想化的智能体的一种刻画, 在形式化实际智能体的过程中, 这些属性并不是我们所期望的, 因为现实中的智能体是受资源约束、能力有限甚至是缺乏意识和演绎能力的实体。 (L04) 对于我们的研究并无任何的坏处, 而且 (L07) 是我们所期望的, 因为这一属性能确保智能体的认知状态是一致的。 (L05) 并非是我们所需要的属性, 它表示一种不受限制的组合。例如, 假设 X 表示智能体的期望 (desire), 智能体期望 φ 和 Ψ 并不意味着智能体期望同时 φ 和 Ψ 。 (L06) 表示不受限制的变弱, 同样亦是我们所不需要的。例如, 智能体期望 φ 并不意味着智能体期望 φ 或 Ψ 。为了克服上述问题, 人们提出了一系列其它方法, 如演绎模型方法、元语言和语法方法^[1,2,4]等等。这些方法用于模型智能体的知识和信念是比较有效的, 因为它能避免逻辑的无所不晓问题。然而它们不适合于描述动机类别和决策类别的认知状态, 因为智能体动机类别和决策类别的认知状态涉及动作概念, 这些方法都是基于断言和命题层次上的, 而要把动作概念引入到这些方法和模型中是比较困难的。文 [7,8,5,9] 提出了一系列组合了动作概念的模态和时态逻辑以描述、规范和分析智能体动机类别和决策类别的认知状态。

尽管正规模态逻辑在智能体理论研究中存在各种各样的不足, 但仍是目前智能体理论研究的主要形式化工具。

3. 研究状况和已有成果

智能体理论研究目前主要涉及三方面的内容: 信息类别, 动机类别和决策类别。三方面的研究是紧密相关的, 尤其是, 动机类别和决策类别的研究与信息类别的研究通常是密不可分的。形式化地研究知识和信念的特征一直是 AI 领域的研究人员所感兴趣的课题。模型知识和信念的核心是确定智能体的信念全集, 亦就是说, 假定智能体相信某个基集 φ 中的每个句子 p , 确定集合 φ' , 使得 $\varphi \subset \varphi'$ 且智能体相信 φ' 中所有句子表达的信息。知识和信念的基于正规模态逻辑的定义如式 (1) 所示。

基于可能世界语义的正规模态逻辑存在逻辑“无所不晓”问题, 因而基于这种方法的知识和信念模型是对智能体的一种理想化解释 (即智能体知道所有的逻辑结果, 知道所有的有效公式等等)。现实的智能体是资源受限, 缺乏意识 (awareness) 和能力有限的实体。为了克服上述问题, 研究人员做了大量的工作。有些工作仍是基于可能世界语义的, 有些工作则采用了其它形式化方法。大致上, 这些方法可分为三类: 语法方法、格局化方法和演绎模型方法。在语法方法中, 智能体的信念由一组句子集来表示。智能体显式地相信某个句子所表达的信息当且仅当该句子出现在智能体的信念集中。语法方法能有效地避免逻辑的“无所不晓”问题, 但缺乏相应的语义理论。用格局化方法来描述智能体的知识和信念的目的是: (1) 对智能体的知识和信念作出语义解释, 同时这种解释又能避免逻辑“无所不晓”问题; (2) 研究和获取智能体的计算特征。在格局化方法中, 定义智能体知识和信念的基本语义实体是格局, 而不是可能世界。可能世界是对世界的一种完全、一致的刻画, 对于每一个原子命题, 可能世界给出它们为真或为假的确切解释, 格局是对世界的部分, 甚至是不一致的认识和刻画。对于某个原子命题, 格局可能将其解释为真或为假或既为真又为假或既不真又不为假。逻辑无所不晓问题的本质是由于可能世界是关于命题的真值函数, 格局的上述模型能有效地克服逻辑的无所不晓问题。这一方面的代表性工作是: Lesvesque 和 Delgrande。Lesvesque 提出了一种隐式和显式信念逻辑。大致思想是: 一个智能体具有一个相对较小的显式信念集和一个相对较大的隐式信念集。Lesvesque 开发了一个具有二个模态算子 (分别表示隐式信念和显式信念) 的逻辑。隐式信念的语义定义仍是基于标准的可能世界语义。显式信念算子的语义定义则基于格局化语义的思想。Konolige 的演绎模型方法将智能体的知识和信念定义为演绎结构 $d = (\Delta, \delta)$ 。其中, Δ 为一基于逻辑语言信念基集, δ 是不完备的推理规则集。Konolige 假定智能体总是尽可能地运用推理规则以生成信念的演绎闭包 $\text{close}((\Delta, \delta)) = \text{def} \{ \varphi \mid \Delta \vdash \neg \neg \varphi \}$ 。智能体信念则由以下式子定义:

$$\text{Bef}_i \varphi \text{ iff } \varphi \in \text{close}(d_i), \text{ 其中, } i \text{ 表示智能体。}$$

动机类别和决策类别的认知状态的研究是当前的研究热点。结合了动作概念的模态或时态逻辑是当前研究的主要形式化工具。然而, 这种方法存在逻辑“无所不晓”问题。许多研究人员正致力于研究如

何对可能世界语义模型作出适当的变更,或采用其他的形式化方法来定义和研究智能体的动机和决策类别的认知状态。这一方面的重要工作之一是 Cohen 和 Levesque 的智能体意愿(intention)理论。Cohen 开展这一方面研究工作的初衷是试图建立言语行为理论,他认为意愿概念是理解言语行为所必须的。Cohen 的工作集中于分析智能体的信念、目标、规划、意愿、承诺等认知状态间所需的理性平衡以及智能体的意愿在维持这种平衡中所起的作用。Cohen^[6]分析了意愿与其他认知状态(目标、期望)间的关系并识别了七种意愿特性以验证理论的有效性。Singh^[7]详细讨论了智能体的意愿性质,并基于一种抽象动作概念(策略)以及命题分枝时态逻辑给出了意愿的定义。Linder 试图开发一个统一的形式化框架来定义和研究动机类别的智能体认知状态。在[9]他采用一种自底向上的方法讨论了智能体的期望、目标和承诺三个认知状态间的关系,并开发了一个形式化的框架定义了这三个认知状态,研究了他们的性质,避免了与逻辑无所不晓相关的副作用问题和传递问题。目前在决策类别认知状态方面的研究相对较少,我们所知道的仅有 Singh 的工作。在[7, 8]Singh 指出:智能体的 BDI 结构不足以描述智能体的理性特征,他认为意愿刻画的仅是智能体对其动作的选择,而期望是一个比意愿更弱的概念。为了描述智能体如何完成动作,必须引入一个新的认知状态以刻画智能体对其动作的控制。基于命题分枝时态逻辑和动态逻辑, Singh 给出了 know-how 的形式化定义,证明了智能体如果具有某个意愿 φ , 智能体持续这种意愿,而且智能体具有相应的能力以及智能体 know-how(φ), 则智能体必然会成功。这一领域的其他工作包括:(1) Moore 的知识和动作理论。Moore 关心的是智能体为了完成动作需要知道哪些信息。他开发了一个包含知识模态子和动作算子的逻辑框架以建立智能体的能力(ability)模型,讨论了具有不完备信息的智能体如何实现目标。(2) Rao 和 Georgeff 的 BDI 体系结构。在研究智能体的体系结构过程中, Rao 和 Georgeff 认为信念、期望和意愿(BDI)是构成智能体的三个重要的认知状态。他们开发了一个分枝时态逻辑以研究这三个概念以及他们之间的各种关系,尤其是,智能体关于将来的信念如何影响它的期望和意愿。(3) Werner 的研究。Werner 的工作利用了经济学、游戏理论、格局自动机理论等等。目前,我们对他的工作还了解不多。

4. 问题和进一步的研究

智能体理论研究是目前正在发展中的一个研究课题,正受到人们日益重视和高度关注。一方面,分布计算是当前影响计算机技术发展和应用的关键技术之一,代表了一种潮流和方向。多智能体系统本质上是一种分布计算系统,同所有分布计算系统一样,它由一组相互交互、通讯、协同和合作的计算实体组成。运用“意向姿态”的观点来研究多智能体系统能帮助我们刻画,理解,分析和预测这类复杂系统的行为,获取、描述、规范和验证系统的需求和性质;另一方面,把智能体视为一种“意向系统”能使我们避开有关智能问题的争论,从一个新的角度去研究和解释智能体的智能特征以及智能体群体间的社会特征(交互,通讯,协同和合作);第三,智能体的理论研究极大地推动了其他相关方法、技术的发展,如 AOP, 基于智能体的软件工程等等,形成了一系列基于智能体概念的方法、技术和工具。尽管智能体的理论研究在许多方面已取得了不少成果,然而仍存在一系列的问题有待人们去解决。首先,现有的研究停留于微观层次(智能体层次)上,即研究构成智能体的各认知状态的概念以及它们之间的关系和性质。然而智能体的社会特征要求我们必须站在更高的层次(社会层次)上开展智能体的理论研究工作,即研究智能体间的交互、通讯、协同和合作。任何智能体都不是独立存在的,它必然要和它所处的环境发生交互。研究、分析和刻画智能体间的交互、通讯、协同和合作,不仅能加强我们对智能体概念的认识,丰富 DAI 的研究内容,促进其他相关计算机学科的发展,而且还能帮助我们描述、分析并最终验证多智能体系统,推动复杂分布系统的开发。目前我们正在开展这一方面的研究工作,并试图将智能体微观层次的研究与宏观层次的研究二者紧密地结合在一起。我们注意到现有的认知类别不足以开展这一方面的研究,至少我们认识到智能体的合作和协商涉及一个与效用值(utility value)相关的认知概念。目前我们对这一概念仍缺乏较深入的认识。其次,目前人们仍不清楚智能体的认知类别间,认知类别的各认知状态间的关系,哪些认知状态是原子的,是构成智能体体系结构对智能体理论研究而言是必须的。尽管在智能体理论研究中,人们已认识到意愿是一个极为重要的概念^[1,6,7,8],但目前仍没有理想的意愿理论模型。第三,形式化工具问题。正规模态和时态逻辑仍是目前智能体研究理论的理想工具,但该方法存在逻辑无所不晓问题以及由此而带来的其他一系列相关问题。此外,开发组合了多个认知状态的逻辑框架

人工智能、面向对象和思维、混沌间的关系

The Relationship between Artificial Intelligence, Object-Oriented and Thinking, Chaos

董 军

(浙江大学 CAD&CG 国家重点实验室 杭州 310027)

摘 要 Artificial intelligence(AI) and object-oriented method(OOM), which have plentiful thinking science intension, are two branches having inseparable relation to thinking. In this paper, taking AI and OOM as centralized aspects, the relationships between AI and OO, OO and thinking, OO and fractal, fractal and chaos, and chaos and AI are described. It is beneficial to interdisciplinary blending and using for reference.

关键词 Artificial intelligence, Object-oriented, Thinking, Cbaos, Fractal

人工智能就是计算机执行某些与人的智能有关的复杂功能的能力^[1], 是对人类智能的一种模拟和扩展。面向对象方法以类、对象、消息和方法为最基本的特征^[2], 追求的是求解空间与问题空间的直接模仿, 即认为数据模型是与概念模型相对应的(数据模型是概念模型在数据世界的抽象描述), 且两种模型各自间的关系也是对应的。人工智能与面向对象是计算机科学中与思维联系紧密的两个分支, 有着丰富的思维科学内涵, 思维科学的基础科学是研究人类意识思维的规律的科学, 即思维学^[3]。智能模拟的核心是思维模拟, 机器对思维模拟的实质在于借

助于原型所拥有的任何特性和功能进行数学描述(模拟)^[4]。人工智能和人类智能的差别除了人脑是自然界长期发展和社会实践的产物以外, 还在于: ①结构不同, 如脑内神经网络并不迭代; ②运动过程不同, 前者是机械、物理、电子的, 后者是生理、心理、反映的。

本文以人工智能和面向对象方法为中心, 阐述人工智能与面向对象、面向对象与思维、面向对象与分形、分形与混沌以及混沌与人工智能之间的关系, 有利于学科间的交融和互相借鉴。

1. 人工智能与面向对象

董军 博士生, 主要从事面向对象方法、分布式人工智能、混沌应用、计算机支持协同工作。

以描述智能体的动态特征和研究智能体性质亦是人们所关心的。最后, 理论研究的目的是推动和促进实践, 如何将智能体的理论研究成果应用于实践, 尤其是应用于面向智能体的程序设计是目前人们关心的研究课题。

参 考 文 献

- [1] Michael Wooldridge et al., Intelligent Agents: Theory and Practice, Appears in Knowledge Engineering Review, 10(2), 1995
- [2] Michael Wooldridge, Time, Knowledge, and Choice. Intelligent Agent: Theories, Architectures, and Languages, Springe-Verlag, 1995
- [3] Afsaneh Haddadi, Kurt Sundermeyer, Acquaintance Relation in Autonomous Agent Societies, 1993
- [4] Joseph Y. Halper, Yoram Moses, A Guide to Comple-

ness and Complexity for Model Logics of Knowledge and Belief, AI, 54, 1992

- [5] Philip R. Cohen, Hector J. Levesque, Intention is Choice with Commitment, AI, 42, 1990
- [6] James P. Delgrands, A Framework for Logics of Explicit Belief. Computational Intelligence, 11, 1995
- [7] Munidar P. Singh, Multiagent System, A Theoretical Framework for Intentions, Know-how, and Communication, Springe-Verlag, 1994
- [8] Munidar P. Singh, Semantical Consideration on Some Primitives for Agents, Same to [2]
- [9] B. vanLinder, et al., Formalizing Motivational Attitudes of Agents, Same to [2]
- [10] Yoav Shoham, Agent-Oriented Programming. AI, 10, 1993