计算机科学 1999Vol 26№ 2

"信念-愿望-意向"Agent 的研究与进展

The Studies and Developments of BDI Agent

路 军 王亚东 王晓龙

(哈尔滨工业大学计算机科学与工程系 哈尔滨 150001)

Abstract This paper discussed the beginning and development of BDI agent in Distributed Artificial Intelligence, analyzed several far-reaching BDI agent model, pointed out their problem and drawback, and then compared their structure each other, gave their logic foundation and application, at last we finger out the direction of BDI agent in the future.

Keywords BDI Agent, Intention, Believe, Object-oriented

1 BDI agent 的起源

在"智能体"(Agent)的研究中引入了许多心理学和人类行为学的概念,一个很有影响的工作来自于 1987 年 Dennett 的"意向系统"(intention system),在这个系统中的行为被描述成诸如"信念"(belief)、"喜好"(preference)和"意向"之类的心智状态,这些状态在决定 Agent 行为时似乎起到了不同的作用。1992 年 Kiss 把它们分为三类:a)认知的(cognitive):如信念和知识等;b)意图的(conative):如意向,承诺和规划等;c)情感的(affective):如愿望,目标和喜好等。

信念、愿望和意向(BDI)通常用来作为这三类心智状态的代表。信念表示一个 Agent 对环境和自身所持的观点,愿望和意图都是 Agent 希望作某事的状态,通常的区别是意向可以作为衡量承诺(commitment)的一个尺度,用来引导和控制 Agent 在未来所作的行为。也就是说:一个 Agent 可能有某种愿望,但有可能永不去履行它;而一旦 Agent 有某种意向,则这种意向将导致 Agent 寻求合适的手段达到这一意向,直到这个意向结束为止。换句话说,意向推动 Agent 去行动,而且意向还对 Agent 未来所作的行为进行引导。这方面的哲学观点来自于 Bratman。

在一些 Agent 结构模型中意向观点被明确表示出来,这些结构模型的一个子集被称作 BDI Agent结构模型。在这些模型里,三元结构(信念,愿望,意向)在 Agent 的认知过程中起到了积极的作用,因此我们称其为 BDI Agent。

2 典型的 BDI Agent 结构模型简介

BDI Agent 最初起源于 Bratman 等人设计的 IRMA 结构及 Georgeff 和 Lansky 设计的 PRS 结构,这两个系统都是针对单 Agent 设计的。而系统中其它 Agent 的存在指明了剩余的问题,相继出现的 COSY 和 GRATE* 的设计则阐明了一些这样的问题。下面我们简要介绍几个典型的 BDI Agent 系统。相关方法和不同方法之间的比较将在第 3 节中给出。

2.1 IRMA

IRMA 为智能的资源受限机器结构(Intelligent Resource Machine Architecture),它是根据"受限资源的理性 Agent"设计的一个实现,这里资源的限制主要指计算能力。

此结构(图 1)由不同的信息库(图中用椭圆表示)和处理过程(用正方形表示)组成,信息库包含信念、愿望,规划库和嵌入规划内的意向。一旦一个意向形成,"手段目的推理器"则被唤醒,分析现存的规划,产生子规划来完成此意向。一个规划一旦变得没有条理的话,"手段目的推理器"就被唤醒,从已存在此意向的规划中提出新的选项作为子规划。

IRMA 理论在 Tileworld 中的一系列试验中被测试和评估,Tileworld 是一个动态和不可预测环境下的一个简单模拟机器人 Agent。试验的主要结果是一个严格的过滤装置再配上一个合适的重叠装置

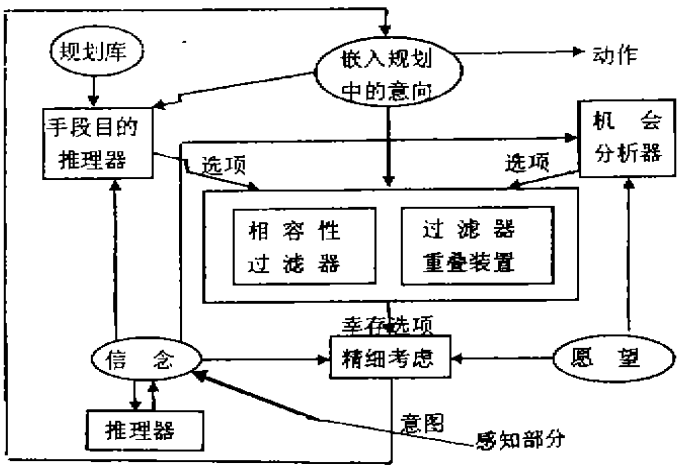

图 1 IRMA 结构(源于 Bratman et al ,1988)

是可行的(至少在某种情况下是可行的);换句话说,对已经采纳的规划的承诺在变化的环境下对一个 Agent 来说是有价值的战略。

2.2 PRS

PRS 是过程推理系统(Procedure Reasoning System)的简称。PRS 是在动态环境下推理和执行任务的一个系统,它是 NASA 的 Space Shuttle 工程中的一个反应控制系统下发展起来的,这个工作基于 Rao 和 Georgeff[6,7]提出的一个完备的理论背景之下。

在本系统中,信念、愿望和意向等观念在任何给定时刻都被明确地表示出来,并合在一起共同决定此系统的动作,这些观念被一个推理机制所推理并且动态地被修改(图 2)。组成这个机制的构件有:数据库,目标,知识域库,意向结构和解释器。

Kinny 和 Georgeff 用 Tileworld 仿真环境来考察对目标的承诺如何影响有效的行为以及对环境的

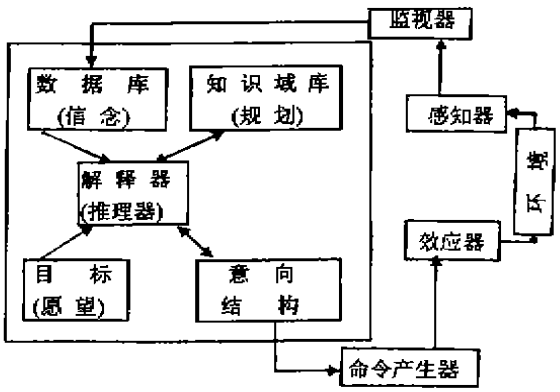

图 2 PRS 结构(源于 Georgeff and Ingrand,1989)

变化作出反应的不同战略间特性的比较。Hanks 等人于 1993 年给出了有用的测试床、经验和有关 Agent 结构的设计。

2.3 COSY

COSY 现在致力于面向 Agent 设计与实现的框架的发展,这个概念已经在诸如交通管理等许多方面得到检验。如图 3 所示,COSY 中 Agent 的结构是定型的(modular),构件包括致动器,感知器,通讯设备,意向和认识结构。前四种的功能不言而喻,认识结构是一个知识库系统,评估感性知识,和意向一起商讨,为给定条件准备一个合适的动作,它包括一个知识库(KB),正本执行构件(SEC),协议执行构件(PEC)和一个"推理、决定和反应"构件(RDRC)。

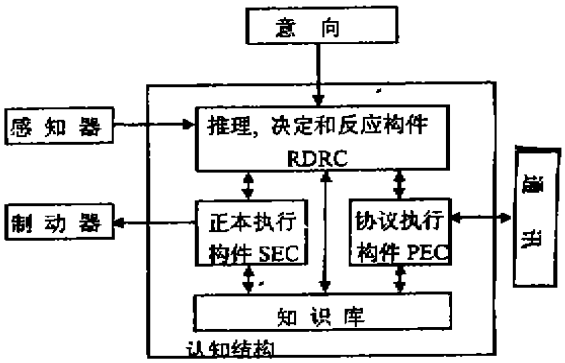

图 3 COSY 的 Agent 结构

三元组(信念,愿望,意向)在 COSY 方法中相当直接地表示出来:KB 包含信念,愿望在意向构件中表示出来(比如战略意向),选择的正本 script 和协议 protocol 代表意向(比如战术意向)。

RDRC 负责推理,并根据需要作出反应。部分任务是处理在相互作用期间产生的决定。它的结构如图 4 所示。

2.4 GRATE*[4]

GRATE* 很像 IRMA,除了带有一个合并合作问题求解的构件,它还使用联合意向(joint intention)和联合责任(joint responsibility)的概念[4]来建立一个合作行为并监视联合行为的执行。在 GRATE* 中第一次将联合意向和联合责任的概念引入到 BDI 结构中,从而使 BDI Agent 的研究进入多 Agent 时代。

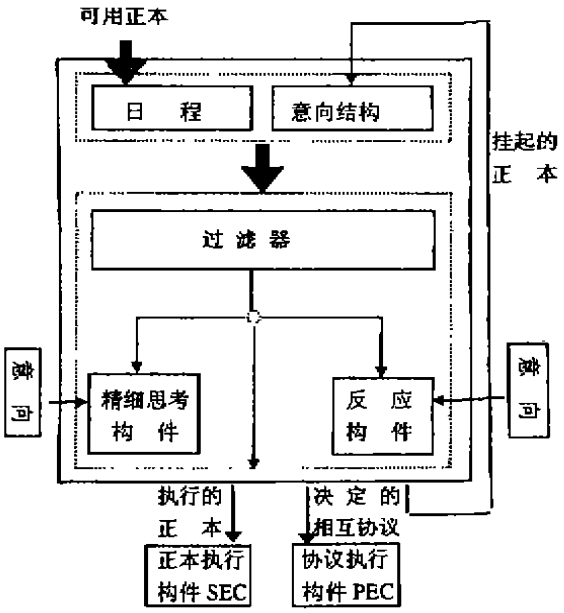

图 4 RDRC 的结构

GRATE* 在电力传输管理上已经被实施并得到验证。如图 5 所示，GRATE* 的结构是一个真正的 BDI 结构：

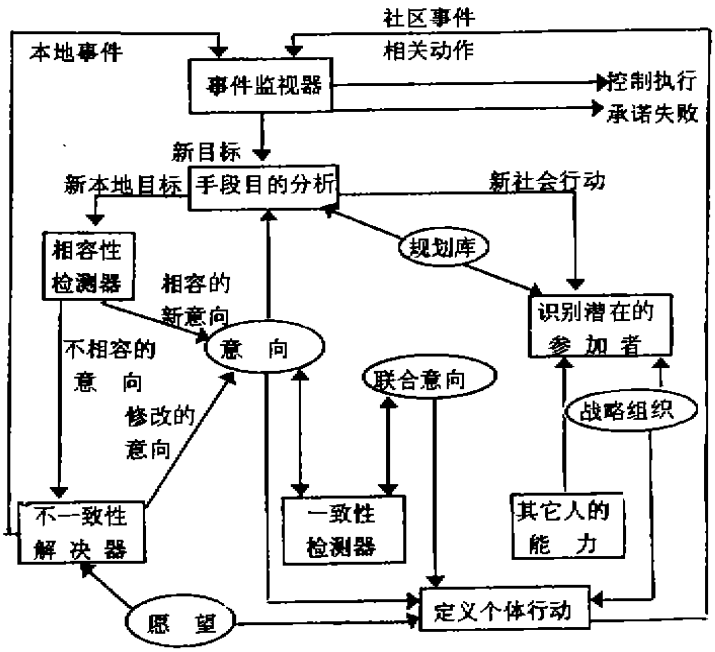

图 5 GRATE* 的功能结构(源于 Jennings，1993)[4]

3 BDI 结构的比较和相关的工作

Bratman 强调信念、愿望和意向是概念上相对独立的三个观念，它们是不可归约的，这就导致了我们在此描述的具有代表性的 BDI Agent 结构的设计。

还有一些有关 BDI Agent 结构刻画的工作，如 hybrid Agent architecture；Chaib-draa 和 Millot 的 Agent 模型；ARCHON；dMARS；INTERRAP；AGENT0，尽管没有直接提到信念、愿望和意向，也引用了许多意向的概念，不过 AGENT0 更强调 Agent 的程序设计而不是它的结构。尽管信念、愿望和意向在 DAI 研究中受到了广泛的关注，但目前为止还没有一个有关其语法和语义的共识，因此我们在对它们进行比较时，更多的只能在概念上和理论上而不是设计上加以比较。

1)IRMA 似乎是 BDI 结构的先驱工作，由于它设计的目标就是计算资源受限的 Agent，所以 IRMA 强调过滤过程和"精细思考装置"的开销在整个系统中的中心地位，它的逻辑基础仅仅是 Bratman 关于意向的哲学分析。

2) 和 IRMA 一样，PRS 是一个单情景(situated)下的 Agent 结构，它也是受 Bratman 的工作影响而设计的，有人发现 PRS 的中心思想可以形式扩展到多 Agent 中去，但目前还没有被应用的报道。在我们描述的 BDI 结构中，PRS 是具有广泛的理论背景的，Rao 和 Georgeff[3] 的逻辑系统给出了 BDI 的语法和语义，在这个最近的工作中[2]，作者改进了目标和愿望在 BDI 结构中的描述：愿望是任务、事务的状态或 Agent 所期望完成的系统行为的集合，这意味着目标堆栈现在和 Agent 的愿望一一对应，目标则被定义为 Agent 选择的愿望，这样，一个 Agent 在给定时间下的意向，是它的目标的一个子集，换言之，是对某个目标的承诺。

3)COSY 则提取了 IRMA 和 PRS 两者的元素，有些模型直接引用这两个结构里的名称，如 intention structure(PRS)，filter，deliberation(IRMA)。其他的构件也可在 IRMA 和 PRS 中清楚

地找到，如 script 对应于 IRMA 规划库中的 plan 和 PRS 中的 KAs。COSY 扩展了它们的这些概念，用合并的合作协议(incorporating cooperation protocols)使 Agent 间的通讯和相互作用成为可能。

4)GRATE* 在很多方面基于 IMRA，是第一个把 BDI 结构应用到联合心智状态上的。

4 BDI Agent 的逻辑描述

在 BDI 方面开展过广泛的理论和逻辑工作，第一个试图形式化这些概念的是 Cohen 和 Levesque[5]，他们以 Bratman 的哲学工作为逻辑基础，其工作为此后发展起来的其他形式化理论奠定了基础。一般地说，所有这些理论都试图通过描述 BDI 之间的相互关系来描述、分析和说明一个 Agent 的行为，这些理论分析的动机千差万别：比如有些是从观察者的角度来解释和预测 Agent 的行为；而另一些则需要设计 Agent 结构的理论基础，甚至从内部描述 Agent 的特点：如社会 Agent。下面我们介绍两个最具代表性的 BDI Agent 的描述。

4.1 Rao 和 Georgeff 的理性 BDI Agent 结构的逻辑描述

Rao 和 Georgeff 用正规模态逻辑(NML)来描述 BDI Agent，在此逻辑中，模态算子的语义是用可能世界上的可达关系来定义的。使用正规模态逻辑的优点是它能很好地描述诸如信念和知识等认知状态，缺点是它不太适合意向的正规理论，会引起"副作用"。

a. 语法：此 BDI 结构的表示语言是计算树逻辑 CTL(Computation Tree Logic)，其中引入可能世界的概念，此逻辑中有两种公式：状态公式(state formulas)和路径公式(path formulas)。

在此逻辑中引入公式 *succeeded*(*e*)和 *failed*(*e*)，分别表示事件 e 刚刚发生的成功或者不成功的行为，*done*(*e*)表示事件 e 刚刚发生的不管是成功还是不成功的行为，*succeeds*(*e*)，*fails*(*e*)，*does*(*e*)具有相似的定义，不过是指将来就要发生的行为。模态词 BEL，GOAL 和 INTEND 分别代表一个 Agent 的信念，目标和意向。

b. 语义：

定义 1 解释器 M 定义为：

$M=\langle W,E,T,\prec,U,B,G,I,\Phi\rangle$

其中：W 是可能世界集；E 是原事件类型集；T 是时间点集；$B\subseteq W\times T\times W$；$G\subseteq W\times T\times W$；$I\subseteq W\times T\times W$；$B_{w,t}^{w}$是信念可达的世界集；$B_{w,t}^{w}=\{b1,b2\}$，表示在时间点 t1，世界 w0 信念可达世界 b1 和 b2，同理可定义 $G_{w,t}^{w}$和 $I_{w,t}^{w}$。

定义 2 可能世界 W 的定义为：

$W=(T_w,A_w,S_w,F_w)$

c. 基本公理：

(AI1)GOAL(α)⊃BEL(α)
(AI2)INTEND(α)⊃GOAL(α)
(AI3)INTEND(does(a))⊃does(a)
(AI4)INTEND(φ)⊃BEL(INTEND(φ))
(AI5)GOAL(φ)⊃BEL(GOAL(φ))
(AI6)INTEND(φ)⊃GOAL(INTEND(φ))
(AI7)done(a)⊃BEL(done(a))
(AI8)INTEND(φ)⊃inevitable◇(¬ INTEND(φ))

4.2 Konolige 和 Pollack 的关于意向的表示主义理论

在这个意向模型中包括两个成分：可能世界(表示事件将来的可能过程)和认知结构(表示一个 Aagent的心智状态成分)。

a. 可能的将来，在此引入两个模态算子：◇和□，分别为可能算子和必然算子，

定义 1 可能算子：$w,W\models\Diamond\phi$ 当且仅当 $\exists\ w'\in W,w',W\models\phi$

必然算子$\Box\phi$定义为$\neg\Diamond\neg\phi$

b. 信念和主意向

定义 2 场景：令 W 为可能世界，ϕ 为 L 中的任意句子，则 ϕ 的场景为集合：

$M_\phi=\{w\in W|w,W\models\phi\}$

定义 3 认知结构：是一个三元组$\langle W,\Sigma,I\rangle$，其中：$W$ 表示可能世界；Σ 是可能世界 W 的一个子集。表示一个 Agent 的信念；I 是 W 上场景的一个子集，表示 Agent 的意向。

c. 理性限制：意向和信念。

d. 相关意向。

5 BDI Agent 的未来发展展望

在 BDI Agent 的研究中，所做工作最多的当属 Rao 和 Georgeff，他们在 1991 年和 1992 年分别给出了 PRS 结构的形式化描述和完整的语义描述，并在此后的几年中不断地更新和发展 BDI 的相关理论和程序设计，于 1995 年在国际人工智能联合会议上提出理性 Agent 意向维护的语义规则[7]，并在 1996 年提出了一种语言 AgentSpeak(L)[8]，为 BDI Agent 理论和应用方面的发展做出了巨大的贡献。在最近的几年里，Kinny 在 BDI 的研究和应用上也表现出浓厚的兴趣，做了许多有益的工作，根据他和

(下转第 47 页)

相关数据的计算，计算模型的有关参数，得到模型的各属性值；通过测试数据对得到的模型进行测试和评价，根据评价结果对模型进行优化。

6）输出结果生成。数据分析的结果一般都比较复杂，很难被人理解，将结果以文档或图表形式表现出来则易于被人接受。

该处理过程模型以用户为中心，通过对用户在进行数据挖掘过程时的工作方式的分析，在设计KDD系统时更注重于对用户的整个数据挖掘的全过程提供支持。

结束语 数据库中的知识发现是一个多阶段的处理过程，作为进行知识提取的数据挖掘在整个过程中虽然起着很大的作用，但其他处理阶段对于知识发现的任务来说也是非常重要的。

处理过程模型描述了KDD中各个处理阶段之间的关系，它对于设计实现以及使用KDD系统都非常重要。目前关于数据库中的知识发现的研究大多局限于数据挖掘即学习算法的研究，这对于它的研究和发展是不利的。我们应该看到数据库中的知识发现是面向实际应用并最终服务于用户，这就需要不仅在学习算法方面进行大量的研究，而且应从总体上对其进行深入的研究，设计出适合实现并易于使用的系统，它们对于KDD的发展是同等重要的。

参考文献

1 王军 数据库知识发现的研究 [中科院软件所博士学位论文] 1997

2 Ram A, Leake D B. A Framework for Goal-Driven Learning. In Proc. of the 1994 AAAI Spring Symposium on Goal-Driven Learning. 1994. 1～11

3 John G H. Enhancements to the Data Mining Process. [Ph. D thesis of Stanford University] 1997

4 Yoon J P, Kerschberg L. A Framework for Knowledge Discovery and Evolution in Databases. George Mason U, ISSE. 1994-July-03

5 Kero B, et al. An Overview of Data Mining Technologies. In The KDD Workshop in the 4th Intl Conf. on *Deductive and Object-Oriented Databases*. Singapore, 1995

6 Chen M -S, et al. Data Mining: An Overview from Database Perspective. IEEE Transactions on Knowledge and Data Engineering, 1997

7 Brachman R J, Anand T. The Process of Knowledge Discovery in Databases: A Human-centered Approach. In: Adavance In Knowledge Discovery And Data Mining. AAAI/MIT Press, 1996

8 Fayyad U M, et al. From Data Mining To Knowledge Discovery. Same to [7]

9 Fayyad U M, et al. Knowledge Discovery and Data Mining, Towards an Unifying Framework. In: Proc. of 2nd Intl Conf. on Knowledge Discovery and Data Mining(KDD-96). AAAI Press, 1996

（上接第51页）

Rao, Georgeff 等人1996年提出的关于BDI Agents系统方法和建模技术[5]，不难预测今后的BDI发展将和面向对象（Object-Orient）技术结合起来，在Agent结构中引入类、对象和实例的概念，并在完善BDI Agents技术的过程中不断扩展面向对象技术。

参考文献

1 Cohen P R, Levesque H J. Intention is choice with commitment. Artif. Intell, 1990. 24: 213～261

2 Georgeff M P, Rao A S. The Semantics of Intention Maintenance for Rational Agents. In: Pro. Int. Jt. Conf. Artif. Intell., 14th, 1995, 704～710

3 Jennings N R. On being responsible. In: Werner E, Demazeau Y, eds. Decentralized Artificial Intelligence. Elsevier/Holland, Amsterdam, 1992. 93～102

4 Jennings N R. Specification and implementation of a belief-desire-joint-intention architecture for collaborative problem solving. Int. J. Intell. Coop. Inf. Syst, 1993, 2(3): 289～318

5 Kinny D, et al. A Methodology and Modelling Technique for Systems of BDI Agents. In: Proc. Eur. Workshop Model. Auton. Agents Multi-Agent World (MAAMAW-96), 7th. 1996. 56～71

6 Rao A S, Georgeff M P. Modeling rational agents within a BDI architecture. In: Allen J, Fikes R, Sandwall E, eds. Proc. of the 2nd Intl. Conf. on Knowledge Representation and Reasoning. Morgan Kaufmann, San Mateo, CA, 1991. 473～484

7 Rao A S, Georgeff M P. An abstract architecture for rational agents. In: Nebel B, Rich C, Swartout W, eds. Proc. of the 2nd Intl. Conf. on Knowledge Representation and Reasoning. Morgan Kaufmann, San Mateo, CA, 1992. 439～449

8 Rao A S. AgentSpeak(L): BDI Agents Speak Out in a Logical Computable Language. In: Proc. Eur. Workshop Model. Auton. Agents Multi-Agent World (MAAMAW-96), 7th. 1996. 42～55