社会 Agent 的思维模型*

何汉明，何华灿
(西北工业大学 计算机学院，陕西 西安 710072)

摘 要：以往对智能 Agent 的社会思维属性的研究，多是将 Agent 的社会性划分到某社会层次进行研究，或者是在个体思维模型中加入某种社会思维状态。然而，智能 Agent 任何时候都是通过承担社会角色而发挥其作用的，而且一旦承担了角色，就完全处于角色的社会环境中，并以角色的方式进行思维，产生角色应有的行为，因此可以认为，智能 Agent 的社会性体现在承担角色的过程中。基于这种思想提出了基于角色的智能 Agent 的社会思维模型，将关系、义务、承诺等社会概念都纳入了角色的范畴之中。

关键词：BDI；角色；基于角色的思维状态；角色承诺；承担角色
中图法分类号：TP18 **文献标识码**：A **文章编号**：1001-3695(2005)07-0026-03

Modeling Mental States of Social Agent

HE Han-ming, HE Hua-can
(College of Computer, Northwest Polytechnicial University, Xi'an Shanxi 710072, China)

Abstract: The research about sociality of Agents usually bases on two ideas: arguing the sociality on social level, or social mental model being joined in BDI. However, an Agent always bring into play his functions by taking on roles, and once he holds a role, he acts as the role by adopting role-based mental attributes in the structure of social relationship of the role. Drawing this idea, the model of role-based social mentality is proposed and some social concepts including relationships, obligations and commitments are hold in the category of roles.

Key words: BDI; Role; Role-based Mental States; Role Commit; Playing Role

1 引言

随着智能 Agent 理论和技术的研究从单个 Agent 扩展到多 Agent 系统，人们发现多 Agent 系统的发展反过来又受到个体行为能力的制约，因此研究智能 Agent 的社会性及智能 Agent 的社会行为能力这一领域受到越来越多的关注。有的研究者认为，研究社会 Agent 及 Agent 的社会性是一种必然趋势，Agent 的社会性将成为多 Agent 系统未来研究和发展的理论基础。

Agent 的社会思维属性是研究 Agent 社会性的一个重要内容，近年来引起了越来越多的重视，并且已经做了不少研究工作[1~7]。Dignum 等人[1]将社会 Agent 分为四个层次：包括知识和信念概念在内的信息层，使用动态逻辑表示动作效果的动作层，包含希望、目标、决策、意图和承诺等概念的动机层和描述了几种 Agent 通信语言的社会层，并用多模态命题逻辑语言 FORM 将它们纳入一个统一的逻辑框架之下，建立起理性社会 Agent 的思维模型。Dignum 等人[2]继而又提出将规范和义务加入 BDI 模型中，以此提高 Agent 推理的有效性。马光伟[3]等人将 MAS 分为社会意识层和个体意识层，认为 BDI 模型中的以意图为中心的观点不适于描述社会 Agent，将意图定义为 Agent 对自身的义务，而将义务看成是 Agent 为他人而怀有的意图，提出以信念、愿望和义务作为基本思维属性的 BDO 模型，以此描述 Agent 的思维状态和社会属性。从另一个角度，Cavedon 等人[4]认为 Agent 承担的角色和角色关系会影响其目标和目标倾向进而影响其行为。Agent 承担角色是对角色目标的承诺，并把角色目标加入到自己的目标偏好结构中，Agent 对角色目标的承诺强度决定了角色目标在目标偏好结构中的位置。Panzarasa 等人[5,6]发展了 Cavedon 的基于角色的社会影响思想，认为社会自然状态明显影响了个体的思维状态，这种影响既可以解释为是承担角色或结合了角色的思维状态，也可以认为是社会和其他 Agent 的思维状态所造成的，因此将这种影响扩展到了包括角色信念、角色目标、角色意图、社会关系等思维属性，形成一个 Agent 的社会心智状态。

可以看出，以上这些工作，一部分是将 Agent 的社会性划分到社会层次进行研究，另一部分则是在个体思维模型中加入社会思维状态。然而，事实上这些思想并不符合人类的思维模式，人类并非只在社会中才进行社会思维，也并非是承担了角色才开始对个体 Agent 的思维状态发生影响，并开始采用社会思维模型。事实上应该是，人类的思维和行为任何时候都是一种社会性的思维和行为，或者说是受到社会的影响，即使是在独自地思维或行为。而这种社会性可以完全纳入角色的范畴之中，体现在承担社会角色之中，一旦承担了角色，就完全处于角色的社会环境中，并以角色的方式进行思维，产生角色应有的行为。因此，可以认为智能 Agent 都是社会中的 Agent，智能 Agent 的作用体现在社会中，进一步说，就是体现在角色中，以角色的身份思维并产生行动。

收稿日期：2004-07-02；修返日期：2004-09-21
基金项目：国家自然科学基金资助项目(60273087)；国家“863”计划资助项目(2002AA412020)

基于以上思想, 通过对智能 Agent 的社会性的分析, 指出智能 Agent 是通过承担角色来发挥它的作用, 而角色都是社会中的角色, 它们必然是处于关系结构之中并发生交互作用。因此描述了包括角色信念、角色目标和角色意图在内的基于角色的思维状态, 而承担角色的智能 Agent 必然以角色的思维状态进行思维和行动。

2　智能 Agent 的社会性

社会学家认为, 社会是由许多个体构成的网所组成的, 这些个体通过各种社会关系互相连接在一起。例如, 大学和图书市场可以看作是整个人类社会中两个不同的小社会, 大学是由校长及其领导下的各职能部门的工作人员、教师和学生以科层式关系组成; 图书市场是由图书销售商、购买商和市场管理员以市场结构组成。这里提到的组成社会的成员包括: 校长、工作人员、教师、学生、销售商、购买商和市场管理员等, 并非指某些具体的个体, 而是指一些角色或角色类型; 使其连接在一起从而形成社会的关系, 也指的是角色关系, 并非指具体哪些个体之间的关系。而一个要参与到社会中的个体, 往往是通过扮演社会中的某个角色加入到角色和角色关系形成的社会结构中, 以符合该角色所内含的一套权利、义务和行为规范并发挥其作用, 从而成为社会的一员。

因此, 可以认为多 Agent 社会是由多个智能 Agent 为了实现各自的或群体的目标, 通过扮演社会中的角色, 以角色关系相互连接而构成的。组成社会的智能 Agent 本身并不具有社会性, 而是在承担社会角色的过程中, 处于与其相关的社会结构中发挥其作用时, 才体现出它的社会性。从另一方面可以认为, 智能 Agent 在发挥功能时, 总是承担了某一角色, 总是以某个角色的身份发挥作用。而如果一个智能 Agent 具有多重功能, 那么它必然是在承担多个不同的角色时发挥这些功能。例如, 教材科某职员的职务是分发教材和购买教材, 分发教材时, 他必然是在学校这个社会结构中, 作为教材分发员的角色来完成这项工作; 而当他购买教材时, 他必然是在图书市场中, 作为购买商的角色去完成教材购买工作; 当他下班回到家里, 又是作为家庭成员做他应该做的事情。因此, 可以认为, 社会 Agent 总是以某个角色的身份, 在该角色所处的社会结构中发挥它的功能。

3　智能 Agent 的 BDI 思维模型

Rao 和 Georgeff 的 BDI 模型[8] 是基于正规模态逻辑 NML 的可能世界模型, 但每个可能世界具有分支时间结构。

一个解释器 M 是一个元组: $M = <W, E, T, ;\ , U, B, G, I, \Phi>$。其中, W 是世界集; E 是原子事件类型集; T 是时间点上的集合; ; 是时间点上的一个二元关系; U 是论域; Φ 是对任何已知的世界和时间点从一阶实体到 U 中元素的一个映射; 关系 B, G, I 将主体当前的处境分别映射到它的信念可达世界、目标可达世界和意图可达世界, 并且 $B \subseteq W \times T \times W$, $G \subseteq W \times T \times W$, $I \subseteq W \times T \times W$。可以用 R 表示这些关系中的任意一个, 并用 R_t^w 表示在时间 t 从世界 w 可达的世界集合。

Bel(a, Φ) 表示主体 a 在时刻 t 具有一个信念 Φ, 当且仅当主体 a 在时刻 t 的所有信念可达世界里 Φ 都为真。

语义为 M, v, w_t　Bel(a, Φ)　*iff* $\forall w' \in B_t^w$, M, v, w'_t　Φ

Goal(a, Φ) 表示主体 a 在时刻 t 具有一个目标 Φ, 当且仅当主体 a 在时刻 t 的所有目标可达世界里 Φ 都为真。

语义为 M, v, w_t　Goal(a, Φ)　*iff* $\forall w' \in G_t^w$, M, v, w'_t　Φ

Int(a, Φ) 用于映射主体当前处境到所有它的意图可达世界。主体 a 在时刻 t 试图使 Φ 为真, 当且仅当 Φ 使主体 a 在时刻 t 的所有意图可达世界里都为真。

语义为 M, v, w_t　Int(a, Φ)　*iff* $\forall w' \in I_t^w$, M, v, w'_t　Φ

Att(a, Φ) 表示主体 a 的思维属性, 即主体 a 在时间 t 具有一个信念(目标或意图) 使 Φ 为真。

4　基于角色的社会思维状态

在社会学中, 角色被认为是构成社会群体或社会组织的基础, 它表示与个体的社会地位相一致的一整套权利、义务和行为规范, 是社会对处在某种特定社会地位的个体的行为期待。对于多 Agent 社会中的角色同样如此。另一方面, 角色是智能 Agent 的抽象描述, 它既不能够推理或决策, 是承担它的智能 Agent 进行推理或决策; 它不会行为, 是承担它的智能 Agent 产生行为。因此角色本身不具有认知和行为能力, 但角色具有承担它的智能 Agent 所应该采取的思维(及行为) 方式, 为了语言上的简便, 称其为基于角色的思维状态, 文中所提到的角色的行为及思维状态都指的是基于角色的行为及思维状态。基于角色的思维状态是指包括角色信念、角色目标和角色意图在内的导致承担它的智能 Agent 行为的思维状态, 而承担角色的智能 Agent 必然以该角色的思维状态进行思维。基于角色的思维状态表示为 RoleAtt(r, Φ) , 即承担角色 r 的 Agent 具有信念(目标或意图) Φ。

如同智能体的信念算子, 角色信念 RoleBel(r, Φ) 表示角色 r 具有信念 Φ。角色信念也可以是不完全的, 但包含了所有可知的角色关于自身、世界和与其相关的其他角色的精神状态的信念, 是承担它的智能 Agent 所必须具有的信念。角色信念包含的与角色相关的内容主要有以下几个方面:

(1) 角色并非孤立存在, 必然处于一个关系结构之中。这种关系结构是对多智能体社会的抽象描述, 被认为是多智能体社会的社会结构。社会结构 SS 可以用一个二元组表示为 $SS = <RO, RE>$, 其中, RO 是社会中所有角色的有限集合; RE 是所有相关的角色之间的关系集合。角色 r 处于角色关系结构 SS 中, 可以用二元谓词表示为 In(r, SS) , $r \in RO$。承担角色 r 的智能 Agent 应该具有关于角色 r 所处的角色关系结构的角色信念, 可以表示为 RoleBel(r, In(r, SS)) 。

(2) 角色具有它相应的义务、权限和规范。义务是角色的职责, 即角色应该履行的责任, 有强制的特性。角色 r 具有义务 O_i, 即角色 r 必须使 O_i 为真, 用二元谓词表示为 Obli(r, O_i) , O_i 是角色 r 的每个职责。权限是角色在履行义务的过程中能够行使的权利, 包括资源使用、访问控制等权限。角色 r 具有权限 P_i, 即角色 r 可以使 P_i 为真, 用二元谓词表示为 Perm(r, P_i) , P_i 代表角色 r 的每个权限。规范是角色的行为规范, 即角色在行使权利和履行义务时必须遵守的规范。角色 r 应该遵守行为规范 N_i, 用二元谓词表示为 Norm(r, N_i) , N_i 是角色 r 的每个行为规范。例如, 学生必须履行其学习义务, 享有

使用学校资源的权利,并且在履行义务和行使权利时,必须遵守学校的各项规章制度。承担角色 r 的智能 Agent 应该具有关于角色 r 必须履行的义务、享有权限和应该遵守的行为规范的角色信念,基于角色的信念可以形式化为 RoleBel(r, Obli(r, O_i) ∧ Perm(r, P_i) ∧ Norm(r, N_i)) 。

(3) 角色关于其所在的关系结构中的其他角色的思维状态的信念也应该是角色信念的一部分,表示为 RoleBel(r, RoleAtt(r', Φ)) , $r \neq r'$。而当时 $r = r'$,表示角色关于自己的思维属性的信念。角色目标是角色被期望达到的目标,为角色的行为指明了方向,是其他角色对该角色的行为应达到的目标的期望。角色目标 RoleGoal(r, Φ) 表示角色 r 具有一个目标使 Φ 为真。角色目标是自己或其他角色对它的期望,因此角色目标也是自己或与其相关的其他角色的角色信念的一部分。角色意图 RoleInt(r, Φ) 表示角色 r 具有意图使 Φ 为真。角色意图是限制角色的可能行为选择,引导和控制角色未来的活动。角色意图还将驱使角色寻求合适的手段达到这一意图。

智能 Agent 通过承担角色成为社会 Agent,可以用二元谓词 Play(a, r) 表示智能 Agent a 承担角色 r。Commit(a, b, r) 表示智能 Agent a 对智能 Agent b 承诺承担角色 r。智能 Agent 承担角色必然意味着智能 Agent 对角色的承诺,因此 Play(a, r) = Commit(a, b, r) 。

智能 Agent 一旦承担角色,就完全以基于角色的思维属性进行思维。因此有 Play(a, r) ∧ RoleAtt(r, Φ) = Att(a, Φ) ,表示智能 Agent a 承担角色 r,并且基于角色 r 的思维属性为 RoleAtt(r, Φ) ,则智能 Agent a 具有的思维属性为 Att(a, Φ) 。

5 结论

以往的模型都是基于单个智能 Agent 的思维状态,难以描述智能 Agent 的社会性思维状态。智能 Agent 的社会性往往是通过承担社会角色,处于角色所在的环境包括角色关系网络之中而体现出来的。因此,智能 Agent 的思维模型应该结合角色,分析在承担角色情况下的思维状态,才有利于其社会思维状态的描述。本文即是从这一思想出发,将关系、义务、承诺等社会概念都纳入了角色的范畴之中,并用基于角色的思维模型描述了智能 Agent 的社会思维状态。进一步的研究工作包括一个智能 Agent 承担多个角色的思维模型,以及角色所处于的关系结构中的其他角色是如何作用于它的思维状态。

参考文献:

[1] ignum F, Van Linder B. Modeling Social Agents: Communication as Action[C]. Proc. of Intelligent Agent, ATAL-96, 1996.

[2] Dignum F, Morley D, Sonenberg E A, *et al*. Toward Socially Sophisticated BDI Agents[C]. Proc. of the 4th International Conference on Multi-Agent Systems, 2000. 118-126.

[3] 马光伟, 徐晋辉, 石纯一. 社会 Agent 的 BDO 模型[J]. 计算机学报, 2001, 24(5): 521-528.

[4] Cavedon L, Sonenberg L. On Social Commitment, Roles and Preferred Goals[C]. Proc. of the 3th International Conference on Multi-Agent Systems, 1998. 80-86.

[5] Panazarasa P, Norman T J, Jennings N R. Modeling Sociality in a BDI Framework[C]. Hong Kong: Proc. of the 1st Asia-Pacific Conf. on Intelligent Agent Technology, 1999. 202-206.

[6] Panazarasa P, Jennings N R, Norman T J. Social Mental Shaping: Modelling the Impact of Sociality on the Mental States of Autonomous Agents[J]. Computational Intelligence, 2001, 17(4): 1-71.

[7] Jennings N R, Campos J R. Towards a Social Level Characterization of Socially Responsible Agents[J]. IEE Proceedings on Software Engineering, 1997, 144(1): 11-25.

[8] Rao A S, Georgeff M P. Modeling Rational Agents within a BDI Architecture[C]. Proc. of the 2th Int. Conf. on Principles of Knowledge Representation and Reasoning, 1991. 473-484.

作者简介:

何汉明,博士研究生,研究方向为人工智能、多智能体系统;何华灿,教授,博士生导师,研究方向为人工智能基础理论、泛逻辑。

(上接第 11 页)

[2] an Goldberg. Privacy-Enhancing Technologies for the Internet II: Five Years Later[C]. LNCS 2482, 2003. 1-12.

[3] Vanja Senicar, Borka Jerman-Blazic, Tomaz Klobucar. Privacy-Enhancing Technologies Approaches and Development [J]. Computer Standards & Interfaces, 2003, 25: 147-158.

[4] D M Kristol. HTTP Cookies: Standards, Privacy, and Politics [J]. ACM Trans. Internet Technology, 2001, 1(2): 151-198.

[5] On-line Privacy Services. http://www.anonymizer.com/, 1999.

[6] David Goldshlag, Michael Reed, Paul Syverson. Onion Routing for Anonymous and Private Internet Connections[J]. Communications of the ACM, 1999, 42(2): 132-143.

[7] M K Reiter, A D Rubin. Crowds: Anonymity for Web[J]. ACM Transactions on Information & System Security, 1998, l(1): 66-92.

[8] M K Reiter, A D Rubin. Anonymous Web Transactions with Crowds [J]. Communications of ACM, 1999, 42(2): 32-48.

[9] D M Goldschlag, M G Reed, P F Syverson. Privacy on the Internet [C]. Proceedings of INET, 1997. 126-131.

[10] P F Syverson, D M Goldschlag, M G Reed. Anonymous Connections and Onion Routiong[C]. Proceedings of IEEE Symousium on Security and Privacy, 1997. 44-54.

[11] L Cranor, M Langheinrich, M Marchiori, *et al*. The Platform for Privacy Preferences 1.0 (P3P 1.0) Specification[R]. W3C Recommendation, 2002.

[12] G Hogben, T Jackson, M Wilikens. A Fully Compliant Research Implementation of the P3P Standard for Privacy Protection: Experiences and Rrecommendations[C]. ESORICS, Lecture Notes in Computer Science 2502, 2002. 104-125.

[13] G Karjoth, M Schunter, E Van Herreweghen. Translating Privacy Practices into Privacy Promises: How to Promise What You can Keep [C]. The 4th Int'l Workshop on Policies for Distributed Systems and Networks, 2003. 211-218.

[14] Gunter Karjoth, Matthias Schunter, Els Van Herreweghen. Amending P3P for Clearer Privacy Promises[C]. Proceedings of the 14th International Workshop on Database and Expert Systems Applications (DEXA '03), 2003. 1529-4188.

[15] T Berners-Lee, J Hendler, O Lassila. The Semantic Web [J]. Scientific American, 2001, 284(5): 34-43.

[16] S McIlraith, T C Son, H Zeng. Semantic Web Services[J]. IEEE Intelligent Systems, 2001, 16(2): 46-53.

[17] A Tumer, A Dogac, H Toroslu. A Semantic Based Privacy Framework for Web Services[C]. Proc. WWW '03 Workshop on E-Services and the Semantic Web, 2003. 356-362.

作者简介:

张军(1966-),男,四川安岳人,副教授,博士,主要研究方向为多媒体信息安全、电子商务安全等;熊枫(1964-),女,重庆人,硕士,研究方向为电子政务。