

Agent

个性化

人工智能

(21)

计算机科学2000Vol. 33No. 3

个性心理学

77-79

个性化 Agent 研究*

The Study of Agent with Personalities

徐晋晖 石纯一

(清华大学计算机系 北京100084)

TP18

Abstract Personality is an important issue in implementing believable and social Agent. This paper summarizes importances of the study, representation of personality and approaches to implementing for Agent with personalities, and presents future work.

Keywords Agent, Personalities

1 概述

Agent 技术是目前人工智能和软件工程的重要研究方面,已得到了广泛应用。行为理性是 Agent 理论研究的核心问题,BDI 逻辑和对策论为逻辑和效用理性的研究提供了较好的工具,但是片面追求最大效用和符合逻辑推论将忽略 Agent 的多样性,会导致单一性的 Agent。随着可信和社会 Agent 的研究,Rizzo 认为当 Agent 赋予个性不同的行为、态度和情感时,用户认为 Agent 是可信的 (believable)^[1],Castelfranchi 认为应该从 Agent 的不同个性组合来理解社会 Agent 的合作行为^[2],以赋予行为理性新的含义,即行为理性指 Agent 的行为是符合个性的。

个性是个性心理学、社会心理学等学科研究的课题,在个性心理学中通常将个性定义为具有一定倾向性的心理特征的总和^[3]。个性结构包括个性倾向性、个性心理特征、心理过程、心理状态和自我调节机制;个性倾向性是指决定人对事物的态度和行为的动力系统,包括需要、动机、兴趣、理想、信念和世界观等不同成分;个性心理特征指在心理活动过程中表现出来的比较稳定的成分,包括能力、气质和性格。心理过程包括认识、情感、意志等心理活动;心理状态是指当前一刻相对稳定的心理背景,是机体受外刺激的影响,在大脑皮层中产生兴奋和抑制活动的独特暂时状态,影响着人所进行的各种活动;自我调节系统以自我意识为核心,在心理结构上包括认识、情感和意志三个方面。个性具有整体性、稳定性、独特性和社会性^[4]。

人工智能领域对个性的研究始于八十年代初 Carbonel 的工作^[1],重点在于人类个性在故事理解中

的作用方面,个性的表示和处理方法为个性 Agent 的研究提供了基础。近两年在 Agent 领域个性的研究成为一个活跃的方向,Castelfranchi 等^[2]分析了多 Agent 合作中委托和帮助者的不同个性对于合作行为的影响,研究了 Agent 个性认识的社会推理,并设计了个性测试床 GOLEM;Rizzo 等^[3]提出了以个性为导向的集思考和反应式规划于一体的 Agent 软件结构;Dryer^[4]以实验的手段分析了 Agent 应该象人一样具有个性在人机交互界面中的重要意义,为个性化 Agent 设计提供了很好的建议,并为 IBM 操作系统 OS/2 设计了体现个性的界面 Agent-WarpGuide。

根据行为=F(个性,情景),可知个性在 Agent 研究中的意义是:(1)Agent 具有个性,可以更好地再现人类智能;(2)Agent 为了获得用户的信任和友善交互,应该具有象人一样的个性,Dryer^[4]通过大量的实现分析说明具有个性的 Agent 可以提高交互的程度;(3)对于环境的认识不仅考虑其他 Agent 的意图和目标,而且要考虑个性,这样才能准确理解和预测环境的变化和其他 Agent 的行为;(4)Agent 建模、分类应该基于个性,例如慷慨的 Agent,自私的 Agent,慷慨和自私是 Agent 的个性;(5)在开放、动态的环境中,没有一个策略是最优的,不同个性的 Agent 具有不同的策略倾向;(6)Agent 的内部状态决定其行为,个性作为一种稳定的内部状态对行为有重要的决定意义。

个性化 Agent 研究应以个性心理学的研究成果为基础,建立对个性可操作的理论和方法。具体来说就是个性化 Agent 模型、个性的表示、结合个性的推理、个性的社会认知、个性的一致性判定、个性的冲突解决以及个性与已有 Agent 研究成果的结合等。

* 本文得到国家自然科学基金和清华大学研究生院博士学位论文基金资助。

目前 Agent 研究中对于个性的理解主要有:(1)个性是 Agent 行为中所表现的一致的、稳定的、可信的、典型的特征和态度集^[1];(2)个性可以表示为有序的目标集和对于目标的规划的偏好程度^[1];(3)个性描述了个体的特殊行为与社会规范行为之间的偏离^[1];(4)个性是在时间和情景上表现一致的个体差异、倾向和性质的集合^[1]。

个性的描述词是很多的,可以分为个性倾向的、个性特征的和 Agent 分类等,在个性 Agent 研究和实现中应准确选择合适的描述词集。Rizzo 等根据帮助行为的特征分析了5种个性特征词:利他的(Altruist)、正常的(Normative)、自私的(Selfish)、同情的(Spiteful)、多疑的(Suspicious);Castelfranchi 等区分了4类委托者:懒惰的(lazy)、依附他人的(hanger-on)、需要时才委托的(delegating-if-needed)、从不委托(never-delegating)、和5类帮助者:超级合作的(hyper-cooperative)、慷慨的(benevolent)、提供的(supplier)、自私的(selfish)、永不帮助的(non-helper);Dryer 建议用友善(Friendliness)和势力(Dominance)度量体系来刻画界面 Agent 的个性。

2 个性表示和实现方法

设 $PS = \{PA_1, PA_2, \dots, PA_m\}$, PA_i 是不同的个性特征描述符。

2.1 标注方法

个性标注 $PM = \{(PA_1, V_1), (PA_2, V_2), \dots, (PA_m, V_m)\}$, V_i 表示倾向度,是一个定量值或定性值(大、较大等)。

每个 Agent 和每个行为均有个性标注,但是含义不同, V_i 分别表示个性特征倾向度,和选择行为的倾向度。个性 Agent 的行为选择是由 Agent 个性标注和行为标注来确定的。

标注方法直观,是一种常见方法,但缺点是个性标注维护困难,缺少个性对目标与策略的倾向性刻画。

Carbone^[1]对于个性的处理模型是标注方法的变形,个性标注描述个性与标准的偏离,即只描述与标准不一样的特征,对应的倾向度表示与标准的相对程度。

2.2 目标方法

目标指的是绝对的、抽象的、高层的目标。例如享乐、社会承认等。既然个性倾向性是以需要为基础的动机系统,那么可以将这里的目标视为需要,而目标方法的基础就是心理学中的需求理论。

设 $WS = \{W_1, W_2, \dots, W_m\}$, W_i 是不同的需求描述符,文[3]中给出的 $WS = \{\text{Resource Provision, Resource Acquisition, Social Responsibility, Belongingness, Image, Entertainment, Safety, Understanding}\}$,

简写为 $WS = \{RP, RA, SP, B, I, E, S, U\}$ 。

个性表示为不同需要的优先程度 $\{(W_1, V_1), (W_2, V_2), \dots, (W_m, V_m)\}$,这反映了不同的个性是由于不同需要决定的。例如对于多疑个性的 Agent 其需要优先程度是 $\{(RP, 0), (RA, 0), (SR, 0), (B, -), (I, 0), (E, 0), (S, +), (U, -)\}$ 。

由于 Agent 个性的不同不仅表现在需要的不同,而且在于实现目标的规划和方法不同。例如要实现“占有一辆汽车”目标,有三种不同的规划:“抢某个 Agent 的”,“与某个 Agent 交换”,“请求某个 Agent 的支助”,不同个性的 Agent 会采取不同的规划。

通常一个规划表示为三元组(前提条件,规划体,结果)。为了能体现不同个性 Agent 对规划的偏好,将规划表示为一个四元组(前提条件,规划体,结果,负作用),负作用描述规划的执行对于 WS 的影响,表示为 $\{(W_1, V_1), (W_2, V_2), \dots, (W_m, V_m)\}$ 。当 $V_i > 0$ 表示对需要加强, $V_i = 0$ 表示对需要无影响, $V_i < 0$ 表示对需要减弱,具体数值表示对应的程度。例如“抢某个 Agent 的”规划的负作用是 $\{(RP, -), (RA, 0), (SR, -), (B, -), (I, 0), (E, 0), (S, 0), (U, 0)\}$ 。

Agent 在实现具体目标的同时要保证不对需要(高层目标)带来负影响,而要产生正影响。基于此,可以给出规划机制中的有关规划的选取和排除知识,根据这些知识指导从大量的可用规划中选取可行规划,可行规划反映了 Agent 的不同个性。

规划排除知识指如果某规划属于可用规划并且规划的结果对于 Agent 的高层的需要有负影响,那么排除该规划。

规划选取知识指如果某规划属于可用规划并且规划的结果对于 Agent 的高层的需要有正影响,那么选取该规划。

基于以上知识可以确定自私的 Agent 选取“抢某个 Agent 的”规划作为实现“占有一辆汽车”目标的规划。

Rizzo 等^[3]基于目标方法,实现了一个集思考型规划器和反应式规划器于一体的 Agent 结构,该结构可以体现个性不同的 Agent 的不同需要层次和实现目标的规划倾向。

2.3 规则方法

Agent 的个性通过一组推理规则来刻画,设 RLS 是推理规则集,那么个性 P 是 PS 到 RLS 的多对多映射,含义就是“如果 Agent A 是 PA,那么 A 采用 P(PA)作为其推理规则”。

例如 A_i 是懒惰的,那么 A_i 可用的委托推理规则有:

(1) $\forall a \forall g[(GOAL, A_i, (T, g)) \text{ AND } (GOAL,$

$$A_i, (EvDoneFor, a, g)) \rightarrow \\ [\exists A_j \neq A_i (BEL, A_i, (CanDo, A_j, a)) \rightarrow \\ (GOAL, A_i, (IntToDo, A_j, a))]$$

设 A_i 有实现 g 的目标和有通过行为 a 完成 g 的目标, 如果存在一个能完成行为 a 的 A_j , 那么 A_i 有使 A_j 完成行为 a 的目标。

$$(2) \{ \forall a \forall g [(GOAL, A_i, (T, g)) \text{ AND } (GOAL, \\ A_i, (EvDoneFor, a, g)) \rightarrow \\ [(NOT \exists A_j \neq A_i (BEL, A_i, (CanDo, A_j, a)) \\ \text{ AND } (BEL, A_i, (CanDo, A_j, a)) \rightarrow \\ (BEL, A_i, (IntToDo, A_j, a))]] \}$$

设 A_i 有实现 g 的目标和有通过行为 a 完成 g 的目标, 如果不存在一个能完成行为 a 的 A_j 和 A_i 相信自己有能力完成 a , 那么 A_i 相信自己意图去完成行为 a 。

$$(3) \{ \forall a \forall g [(GOAL, A_i, (T, g)) \text{ AND } (GOAL, \\ A_i, (EvDoneFor, a, g)) \rightarrow \\ [(NOT \exists A_j \neq A_i (BEL, A_i, (CanDo, A_j, a)) \\ \text{ AND } (BEL, A_i, NOT(CanDo, A_i, a)) \rightarrow \\ (BEL, A_i, (CurrentlyUnachievable, a, g))]] \}$$

设 A_i 有实现 g 的目标和有通过行为 a 完成 g 的目标, 如果不存在一个能完成行为 a 的 A_j 和 A_i 相信自己无能完成 a , 那么 A_i 相信目前不能完成行为 a 来实现目标 g 。

Rao 和 Georgeff^[5]建立的三种不同个性的 Agent 所采纳的承诺机制, 就是规则方法; Castelfranchi 等^[2]实现的多 Agent 合作测试床 GOLEM 中个性的实现也是基于规则的方法。

2.4 个性推导

在个性—目标—规划—行为之间存在一个环, 正向环是指个性决定目标, 而目标的实现需要规划, 规划的结果触发行为, 反向环是从行为可以识别规划, 根据规划可以确定目标, 目标和规划可以确定个性。

上面的方法基本描述了正向环, Agent 为了有效地与其它的 Agent (或人) 交互, 必须建立起其他 Agent 的模型, 现有的关于其他 Agent 的推导主要在于信念、目标、意图、能力的认识, 缺少对于个性的认识机制。

根据个性心理学中的归因原理可以建立个性的推导机制, 文[2]中给出了一些实例化的推导规则。

例: 如果 A_i 请求 A_j 完成 a , 那么 A_i 可以认为 A_j 是懒惰的或依靠他人的或需要时才委托的; 如果 A_j 拒绝 A_i 的请求, 那么 A_i 可以认为 A_j 不是超级合作的。

结束语 由于个性化 Agent 研究是一个崭新的课题, 尚有许多问题需要解决, 针对现有的研究工作, 这里给出进一步有意义的工作方向。

(1) 建立形式化的个性 Agent 模型。现有个性 Agent 研究处于实现层次, 没有形式化的理论模型作指导。

(2) 建立有效的个性表示语言。由于个性的描述词非常多, 应该分析个性不同的核心成分, 以核心成分为度量体系来描述不同的个性。

(3) 与现有 Agent 研究的结合。BDI 逻辑和对策论是目前 Agent 理论研究的主要方法, 应该建立集个性理论, BDI 逻辑和对策论于一体的 Agent 结构。

(4) 个性的冲突和一致性判断和解决。Agent 可以有多个个性, 在多 Agent 系统中应该提供个性冲突的解决方法。

(5) 灵活的个性机制。个性是稳定的, 但不是一成不变的, 随着与其他 Agent 的交互, 个性可以进行调节, 以提高在 Agent 社会中的生存能力。

参考文献

- 1 Carbonel J G. Towards a process model of human personality traits. *Artificial Intelligence*, 1980, 15(2): 49~74
- 2 Castelfranchi C, et al. Personality traits and social attitudes in multiagent cooperation. *Applied Artificial Intelligence*, 1998, 12(7-8): 649~675
- 3 Rizzo P, et al. Goal-based personalities and social behaviors in believable agents. *Applied Artificial Intelligence*, 1999, 13(2): 239~271
- 4 Dryer D C. Getting personal with computers: how to design personalities for agents. *Applied Artificial Intelligence*, 1999, 13(2): 273~295
- 5 Rao A S, Georgeff M P. Modeling rational Agents within a BDI-architecture. In: Allen J, Fikes R, Sandewall W, eds. *Principles of Knowledge Representation and Reasoning*. California: Morgan Kaufmann Publishers, 1991. 473~484
- 6 叶奕乾, 孔克勤. 个性心理学. 华东师范大学出版社, 1993
- 7 朱智贤主编. 心理学大词典. 北京师范大学出版社, 1989