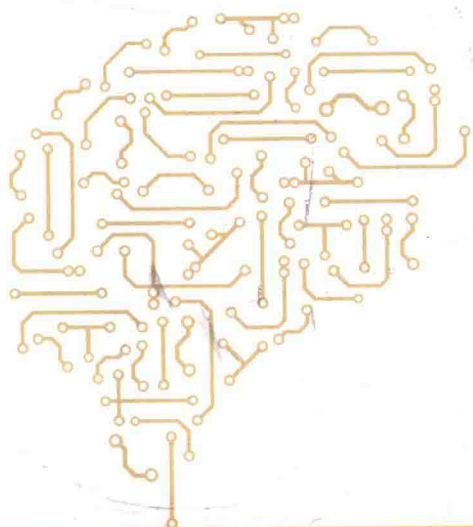


MULTIAGENT SYSTEMS
AND THEIR APPLICATIONS (VOLUME 2)

多智能体系统及应用

王崇骏 史忠植 常亮 王文剑 主编

卷二



清华大学出版社



多智能体系统及应用 卷二

MULTIAGENT SYSTEMS
AND THEIR APPLICATIONS
(VOLUME 2)

王崇骏 史忠植 常亮 王文剑 主 编

清华大学出版社
北 京

内 容 简 介

智能体与多智能体系统是人工智能领域活跃的研究分支,其涉及的基本研究问题包括合作协调、推理规划、学习、工具框架、模型等。本书的内容选自2015年及2016年度“中国智能体及多智能体系统”研讨会的特邀报告,同时收集了部分学者的最新研究进展。全书共计11个专题,每个专题均是智能体领域的重点及热点问题,分别涉及:多智能体学习、多智能体协调协商、多智能体规划、多智能体应用框架及工具以及交叉应用基础问题研究。

本书可供高等院校、科研院所计算机、自动化及相关专业的师生、科技工作者及相关企业的工程技术人员阅读参考。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

多智能体系统及应用. 卷二/王崇骏等主编. —北京:清华大学出版社,2018

ISBN 978-7-302-48777-7

I. ①多… II. ①王… III. ①人工智能—研究 IV. ①TP18

中国版本图书馆CIP数据核字(2017)第272821号

责任编辑:薛 慧

封面设计:常雪影

责任校对:王淑云

责任印制:王静怡

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦A座 邮 编:100084

社总机:010-62770175 邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质量反馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

印装者:三河市君旺印务有限公司

经 销:全国新华书店

开 本:185mm×260mm 印 张:13.5 彩插:2 字 数:333千字

版 次:2018年1月第1版 印 次:2018年1月第1次印刷

印 数:1~2000

定 价:58.00元

产品编号:076579-01

前言

智能体(agent)与多智能体系统(multi-agent systems)是人工智能领域一个活跃的研究分支。20世纪90年代,随着计算机网络、计算机通信等技术的发展,对于智能体与多智能体系统的研究不仅成为分布式人工智能研究的一个热点,而且也成为信息技术关注的一个热点。智能体是一种处于一定环境下的系统,它能在那种环境下灵活、自主地活动。智能体提供了一种新的计算和问题求解风范。人们在研究人类智能行为中发现:人类绝大部分的活动都涉及由多个人构成的社会群体,大型复杂问题的求解需要多个专业人员或组织协作完成。最重要的和最复杂的智能是在由众多个体构成的社会中进行各种活动时体现出来的。“协作”“竞争”“谈判”“博弈”等是人类智能行为的重要表现形式。智能体与多智能体系统的研究与逻辑学、运筹学、经济学、博弈论、社会学等学科均有密切的联系,有着精彩、深厚的理论积淀。

近十多年来,智能体和多智能体技术在世界上获得了广泛应用,如AlphaGo、德州扑克、星际争霸(StarCraft)、无人机协同、反恐等。国际学术界陆续创办了专门面向智能体及多智能体系统研究的国际学术期刊 *Autonomous Agents and Multi-Agent Systems* 和国际学术会议“The International Conference on Autonomous Agents and Multi-Agent Systems”。除此而外,在人工智能领域的主要国际学术期刊(如 *Artificial Intelligence*, *Journal of Artificial Intelligence Research*)和会议(如IJCAI和AAAI)上,关于智能体与多智能体系统的研究论文也占据了非常高的比例。

2006年,中国计算机学会人工智能与模式识别专业委员会发起组织了两年一度的“全国Agent理论与应用学术会议”,并从2006年至2012年分别在烟台大学、南京大学、国防科技大学和吉林大学一共举办了四届。2013年起,该会议并入人工智能与模式识别专业委员会在当年开始举办的两年一度的“中国计算机学会人工智能会议(CCF-AI)”。为了更好地推动中国智能体和多智能体系统理论、技术和应用的发展,2014年3月由史忠植研究员、安波副研究员主持,在中国科学院计算技术研究所举办了首届“中国智能体及多智能体系统”研讨会,该会议采用了不征文、不收费、报告人由组织者邀请的新方式。在首届会议中,与会者一致同意在中国计算机学会人工智能与模式识别专业委员会指导下,申请成立“多智能体与智能系统学组”,决定每年由该学组组织一次“中国智能体及多智能体系统研讨会”,并对邀请的报告进一步扩充、润色后结集出版。

2015年,由高阳教授、安波副研究员、陈小平教授、毛新军教授主编的《多智能体系统与应用》就是在2014年度的部分特邀报告基础上结集出版的。

本书的内容遴选自2015年、2016年度中国智能体及多智能体系统研讨会的特邀报

告并收集了部分学者的最新研究进展。全书共计 11 个专题,每个专题讨论一个技术主题或者应用,并兼顾该主题的综述和特邀报告人自身的研究成果。书中章节涉及多智能体学习、多智能体协调协商、多智能体规划、多智能体应用框架及工具以及交叉应用基础问题研究。

本书的出版得到了中国计算机学会人工智能与模式识别专业委员会的支持和指导,在此表示衷心的感谢。本书的编写和出版还得益于各个章节的各位作者的无私支持和帮助,在此表示特别的感谢。中国科学院计算研究所的安波副研究员、南京大学的高阳教授、燕山大学的张大鹏教授、山东科技大学的纪淑娟副教授为本书的出版提出了很多的宝贵意见,再次对他们表示感谢。还要感谢南京大学的吴骏博士、张雷博士以及宋岳、许磊、唐思雨、乔羽、陈港、李博、叶康、经纬等同学,在本书编撰及整理润色的过程中,他们给予了极大的帮助。

由于水平有限以及知识观及价值观的狭隘,书中的疏漏和不足之处在所难免,敬请各位专家和读者批评指正。

编 者

2017 年 5 月

目 录

专题 1 智能体动机学习	史忠植, 马 刚, 李建清(1)
1 概述	(1)
2 动机理论	(2)
3 动机学习	(6)
4 基于动机的强化学习	(9)
5 小结	(15)
参考文献	(15)
专题 2 多智能体强化学习中的博弈、均衡和知识迁移	胡裕靖, 高 阳(17)
1 绪论	(17)
2 背景知识	(20)
3 不共享值函数的多智能体强化学习	(24)
4 基于均衡迁移的多智能体强化学习	(31)
5 稀疏交互的多智能体系统中的知识迁移和博弈约简	(37)
6 小结	(44)
参考文献	(45)
专题 3 一种基于中介 agent 的强化学习优化协商方法	陈利红, 董红斌(48)
1 引言	(48)
2 基于强化学习的双边多议题优化协商模型	(49)
3 基于中介 agent 自适应学习的协调协商	(53)
4 实验设计与分析	(55)
5 小结	(64)
参考文献	(64)
专题 4 多智能体协调技术综述	郝建业(66)
1 引言	(66)
2 固定对手重复博弈	(68)
3 合作式群体博弈	(72)

4 非合作式群体博弈	(74)
5 小结	(77)
参考文献	(77)

专题 5 基于决策理论的多智能体规划 吴 锋(81)

1 引言	(81)
2 分布式局部可观察马尔可夫决策过程(DEC-POMDP)	(83)
3 DEC-POMDP 的离线规划算法	(85)
4 DEC-POMDP 的在线规划算法	(90)
5 小结	(92)
参考文献	(93)

专题 6 部分可观察环境中的序贯决策理论及方法研究 章宗长(94)

1 引言	(94)
2 部分可观察马尔可夫决策过程	(95)
3 离线规划方法	(99)
4 在线规划方法	(104)
5 基于覆盖数的规划理论	(107)
6 小结	(108)
参考文献	(109)

专题 7 基于 Landmark 的启发式搜索规划 张 雷, 吴 骏, 王崇骏(112)

1 引言	(112)
2 STRIPS 规划问题	(112)
3 STRIPS 问题的启发式函数设计	(114)
4 多值 Landmark 规划	(118)
5 小结	(127)
参考文献	(127)

专题 8 AutoRobot: 基于多主体系统的自主机器人软件框架

..... 毛新军, 杨 硕, 杨 森(129)

1 引言	(129)
2 自主机器人及其软件特点	(131)
3 机器人软件技术分析	(135)
4 基于多主体系统的自主机器人软件体系结构	(139)
5 自主机器人软件框架 AutoRobot	(144)
6 案例分析	(147)
7 小结	(151)
参考文献	(151)

专题 9 计算经济学与最优机制设计问题	唐平中(154)
1 计算经济学简介	(154)
2 最优机制设计问题:单件商品	(155)
3 最优机制设计:多件商品	(159)
4 小结	(160)
参考文献	(160)
专题 10 基于计算博弈论的出租车服务定价研究	甘家瑞,安 波(162)
1 研究背景与相关工作	(162)
2 博弈建模	(164)
3 模型求解——紧凑表达法	(169)
4 ASM 算法——基于元时间表的紧凑表达	(169)
5 解决任意约束下的问题	(179)
6 实验	(183)
7 小结	(187)
参考文献	(189)
专题 11 云计算中定价机制的研究	秦 涛(191)
1 云计算简介	(191)
2 云计算中的定价模型	(193)
3 市场竞争、演变与定价策略优化	(196)
4 小结	(205)
参考文献	(206)

专题 1 智能体动机学习

史忠植, 马 刚, 李建清

中国科学院计算技术研究所智能信息处理重点实验室, 北京 100190

摘 要 动机是直接驱动智能体行为以达到一定目的的内在动力和主观原因。动机为激活、引导和维护智能体行为随着时间推移的内部过程。动机触发多智能体协同工作。本文提出了一种基于环境感知的动机学习算法, 也讨论了基于动机的强化学习方法。

关键词 智能体; 强化学习

1 概述

动机(motivation)是直接驱动智能体行为以达到一定目的的内在动力和主观原因。动机与激活、引导和维护的行为一样都是随着时间变化的内部过程。在文献[1]中穆克(D. G. Mook)简单地定义动机是“行动的起因”。

1943年, 马斯洛(A. H. Maslow)提出动机的需求理论^[2]。马斯洛假定, 人的需要, 即人的动机顺序发生, 从最基础的生理和安全的需要, 通过一系列的爱和尊重的需要, 发展为自我实现的复杂需求, 而需要层次有着巨大的直观吸引力^[5]。多年来, 人们提出许多动机理论, 每种理论都在某种程度上有着不同的关注点。这些理论尽管在许多方面十分不同, 但它们都出自相似的考虑, 即对行为的唤起、指向和维持, 这三点是任何一种动机分析的核心。

格林(R. G. Green)等人将动机理论分为生理、行为和社会的3类^[3]。梅里克(K. E. Merrick)将动机理论分为4大类, 即生物学理论、认知理论、社会理论和组合动机理论^[4]。生物学动机理论试图依据自然体系生物学层面的工作过程解释动机。这些理论的机理经常采用能量和运动方式解释行为, 使得生物体朝向一定行为。现有的人工系统研究已经使用生物学动机理论创建软件智能体和进行自然系统的模拟。

饥饿和口渴可被看作体内驱动的运动或者标志最佳的唤醒理论, 意味着吃喝或者探查是生理状态监控变化被起动。不过, 除发生响应生理的变化之外, 类似馈送和喝水的行为也与这种体内运动有关。由此可见, 认知动力理论集中于怎样确定行为, 结果怎样影响行为和影响到什么程度, 根据不同的行动步骤的费用和效益, 解释个人行为将来很可能的结果。基于抽象的机器学习和人工智能概念, 例如目标、规划、策略, 动机的认知理论可以为动机计算模型提供一个初始点。

社会动机理论涉及个体与他人接触过程中的行为。动机的社会理论是生物学和认知理论的交叉。例如采用适合度和文化效应描述认知现象, 而进化论可以被认为是生物学社会理论。社会动机理论可以从小组态势下的个人到更大的社会、文化和进化系统。这些理论为多智能体系统动机计算模型的设计提供重要的初始状态。

组合动机理论尝试综合生物学、认知和社会动机理论,例如,马斯洛的需求层次学说^[5]、奥尔德弗的 ERG 理论^[6]以及斯塔格纳的稳态模型^[7]。对于人工系统动机综合模型也是研究的重点,这种模型在硬件、抽象推理和多智能体层面提供描述行为过程的综合算法。

人的各种行为和活动都离不开动机,动机有下列功能:

- 唤起行动的起动机。就个人来说,他的行动的一切动力,都一定要通过他的头脑,一定要转变为他的愿望的动机,才能使他行动起来。
- 维持活动达到目标的志向功能。动机一旦引起行为和活动,并能使这种活动具有稳固而完整的内容,使人表现出极大的积极性,朝思暮想,茶饭不香,思维敏捷,能持久而顽强地进行这种活动。
- 动机的强化功能。一个人在活动上的成功和失败的体验,对他的活动志向有一定的影响。或者说,行为的结果如何,影响着人的动机。由此可知,动机对人的行为起着以正负强化形式出现的调节控制作用。

2 动机理论

动机研究大致上可以分为三个阶段:20 世纪 60 年代之前,主要以行为主义和精神分析理论为主导,强调本能、冲动、驱力、体内平衡等生物性的因素在决定人的动机和行为方面的直接作用。20 世纪 60 年代以后,认知的观点逐步介入到动机研究中来,研究的课题发生了很大的变化,出现了归因理论等强调认知因素的动机理论,并使传统的基于行为主义观点的自我效能理论、习得无助理论在内容上发生了巨大的变化。20 世纪 80 年代后的动机研究,已经逐步走向整合,构建一个具有普遍意义的动机理论。下面重点讨论几种动机理论。

2.1 需要层次理论

人本主义心理学家马斯洛试图统一大量人类的动机研究成果,提出需要层次理论^[5],如图 1 所示。他致力于对人的动机研究,认为人有 5 种基本的需要,按其满足的先后依次排列成一个层次。在这一层次中,最基础的生理方面的需要,即对食物、水、空气等的需要;在生理需要得到基本满足之后,便出现安全或保护的需要;随后出现对爱、感情、归属的需要;接着出现对尊重、价值或自尊的需要;在上述这些低一级的需要得到基本满足之后,最后剩下的便是对自我实现的需要。所谓自我实现,就是使自己更完备、更完美,能够更充分地使用自己具有的能力和技能。马斯洛认为,人的绝大部分时间和精力都用于旨在实现最基本的但又尚未满足的需要上,当这些需要或多或少得以实现后,人才能越来越注意到更高层次的需要。他认为,在这些需要中,前四种是缺失性需要,它们对生理和心理的健康是很重要的,必须得到一定程度的满足,但一旦得到满足时,由此而产生的动机就会消失。最后一种需要即自我实现的需要,它是成长需要,很少得到完全的满足。而对一个正常健康的人来说,因缺失性需要已得到相当的满足,所以他们的行为是由不同类型的成长需要所决定的。需要层次理论对临床和咨询心理产生了影响,并成为其动机理论的基础。



图 1 马斯洛的层次需要模型

在马斯洛的动机需求理论基础上,巴赫(J. Bach)提出了一种认知智能体的扩展动机框架^[8]。他描述了与系统需求相关的预定义的有限的驱动集。目标通过与环境交互的强化学习来达到。巴赫同时指出 Psi 智能体的所有行为都直接指向目标状态。目标状态为满足所有需求的消费行为。巴赫提出了智能体需求的三层理论。最底层为生理需求,包括燃料,水,完整性。第 2 层为认知需求,包括确定性、完整性和审美。第 3 层为社会需求,包括加入和请求信号。巴赫使用 Psi 理论来对一个基于动机的、多智能体的动机系统的可能求解。它既能反应出生理需求,也能表达认知和社会需求。它直接集成了需求,能够快速适应不同的环境和智能体。他们开发了一种模型能成功评估人类在进行问题求解游戏时的表现。

2.2 好奇心成长理论

成长理论的一个基本概念是人类并非生来就有完全发展的能力。为了成功地与环境打交道,需要尽可能地向环境学习并最大限度地发展他们的技能。桑德斯(R. Saunders)和格罗(J. S. Gero)^[9,10]借鉴了马斯兰(S. Marsland)的工作^[11],基于新颖性,开发了好奇心和兴趣计算模型。他们使用实时新奇检测器来发现新颖性。桑德斯和格罗也采用组合神经网络表示新颖性,利用强化学习创建好奇设计智能体。

1997 年,施密德胡贝尔(J. Schmidhuber)创建了一个新奇的、创造性的探险家,具有两个共同进化的大脑作为兴趣模型^[12]。欧德耶(P.-Y. Oudeyer)和卡普兰(F. Kaplan)等开发的智能自适应好奇系统(IAC)作为机器人的动机系统^[13,14],鼓励关注情景能最大限度地提高学习的进步。他们的模型 IAC 朝向维护驱动,使抽象的动态认知变量学习进步保持最大。他们称其为好奇认知模型,推动机器人智能体可以学习抽象的情景表示。

2.3 成就动机理论

美国哈佛大学麦克利兰(D. C. McClelland)从 20 世纪 40~50 年代开始对人的需要和动机进行研究,提出了著名的“三种需要理论”,他认为个体在工作情境中有三种重要的动机或需要:成就需要、权力需要、亲和需要^[15]。麦克利兰认为,具有强烈的成就需求的

人渴望将事情做得更为完美,提高工作效率,获得更大的成功,他们追求的是在争取成功的过程中克服困难、解决难题、努力奋斗的乐趣,以及成功之后的个人的成就感,他们并不看重成功所带来的物质奖励。权力需求是指影响和控制别人的一种愿望或驱动力。不同人对权力的渴望程度也有所不同。权力需求较高的人对影响和控制别人表现出很大的兴趣,喜欢对别人“发号施令”,注重争取地位和影响力。他们常常表现出喜欢争辩、健谈、直率和头脑冷静;善于提出问题和要求;喜欢教训别人并乐于演讲。亲和需求就是寻求被他人喜爱和接纳的一种愿望。高亲和动机的人更倾向于与他人进行交往,至少是为他人着想,这种交往会给他带来愉快。高亲和需求者渴望亲和,喜欢合作而不是竞争的工作环境,希望彼此之间的沟通与理解,他们对环境中的人际关系更为敏感。

阿特金森(J. W. Atkinson)的成就动机理论被认为是一种期望价值理论,因为这一理论认为动机水平依赖于一个人对目的的评价以及达到目的可能性的评估^[16]。阿特金森重视冲突的作用,尤其重视成就动机与害怕失败之间的冲突。该理论的特征是它可以用数量化的形式来说明。阿特金森认为,最初的高成就动机来源于孩子生活的家庭或文化群体,特别是幼儿期的教育和训练的影响。个人的成就动机可以分成两部分:其一是力求成功的意向;其二是避免失败的意向。也就是说,成就动机涉及到对成功的期望和对失败的担心两者之间的情绪冲突。

2.4 激励理论

激励理论是关于如何满足人的各种需要、调动人的积极性的原则和方法的概括总结。激励的目的在于激发人的正确行为动机,调动人的积极性和创造性,以充分发挥人的智力效应,做出最大成绩。自从20世纪20~30年代以来,国外许多管理学家、心理学家和社会学家结合现代管理的实践,提出了许多激励理论。这些理论按照形成时间及其所研究的侧面不同,可分为行为主义激励理论、认知激励理论和综合型激励理论3大类。

(1) 行为主义激励理论。20世纪20年代,美国风行一种行为主义的心理学理论,认为管理过程的实质是激励,通过激励手段,诱发人的行为。在“刺激—反应”这种理论的指导下,激励者的任务就是去选择一套适当的刺激,即激励手段,以引起被激励者相应的反应标准和定型的活动。

新行为主义者斯金纳在后来又提出了操作性条件反射理论。这个理论认为,激励人的主要手段不能仅仅靠刺激变量,还要考虑到中间变量,即人的主观因素的存在。具体说来,在激励手段中除了考虑金钱这一刺激因素外,还要考虑到劳动者的主观因素的需要。根据新行为主义理论,激励手段的内容应从社会心理观点出发,深入分析人们的物质需要和精神需要,并使个体需要的满足与组织目标的实现一致化。

新行为主义理论强调,人们的行为不仅取决于刺激的感知,而且也决定于行为的结果。当行为的结果有利于个人时,这种行为就会重复出现而起着强化激励作用。如果行为的结果对个人不利,这一行为就会削弱或消失。所以在教育中运用肯定、表扬、奖赏或否定、批评、惩罚等强化手段,可以对学习者的行为进行定向控制或改变,以引导到预期的最佳状态。

(2) 认知激励理论。行为简单地看成人的神经系统对客观刺激的机械反应,这不符

合人的心理活动的客观规律性。对于人的行为的发生和发展,要充分考虑到人的内在因素,诸如思想意识、兴趣、价值和需要等。因此,这些理论都着重研究人的需要的内容和结构,以及如何推动人们的行为。

认知激励理论还强调,激励的目的是要把消极行为转化为积极行为,以达到组织的预定目标,取得更好的效益。因此,在激励过程中还应该重点研究如何改造和转化人的行为。属于这一类型的理论还有斯金纳的操作条件反射理论和挫折理论等。这些理论认为,人的行为是外部环境刺激和内部思想认识相互作用的结果。所以,只有改变外部环境刺激与改变内部思想认识相结合,才能达到改变人的行为的目的。

(3) 综合型激励理论。行为主义激励理论强调外在激励的重要性,而认知激励理论强调的是内在激励的重要性。综合型激励理论则是这两类理论的综合、概括和发展,它为解决调动人的积极性问题指出了更为有效的途径。

心理学家提出的场动力理论是最早期的综合型激励理论。这个理论强调,对于人的行为发展来说,先是个人与环境相互作用的结果。外界环境的刺激实际上只是一种导火线,而人的需要则是一种内部的驱动力,人的行为方向决定于内部系统的需要的强度与外部引线之间的相互关系。如果内部需要不强烈,那么,再强的引线也没有多大的意义。

波特(L. Porter)和劳勒(Edward E. Lawler)于1968年提出了新的综合型激励模式,将行为主义的外在激励和认知的内在激励综合起来^[17]。在这个模式中含有努力、绩效、个体品质和能力、个体知觉、内部激励、外部激励和满足等变量。波特与劳勒把激励过程看成外部刺激、个体内部条件、行为表现、行为结果相互作用的统一过程。一般人认为,有了满足才有绩效。而他们则强调,先有绩效才能获得满足,奖励是以绩效为前提的,人们对绩效与奖励的满足程度反过来又影响以后的激励价值。人们对某一作业的努力程度,是由完成该作业时所获得的激励价值和个人感到做出努力后可能获得奖励的期望概率所决定的。很显然,对个体的激励价值越高,其期望概率越高,则他完成作业的努力程度也愈大。同时,人们活动的结果既依赖于个人的努力程度,也依赖于个体的品质、能力以及个体对自己工作作用的知觉。

2.5 归因理论

归因理论是指说明和分析人们活动因果关系的理论,人们用它来解释、控制和预测相关的环境,以及随这种环境而出现的行为,通过改变人们的自我感觉、自我认识来改变和调整人的行为的理论。1958年,奥地利社会心理学家海德(F. Heider)在他的著作《人际关系心理学》中,首先提出归因理论^[18],从通俗心理学(naive psychology)的角度提出了归因理论,该理论主要解决的是日常生活中人们如何找出事件的原因。海德用归因理论发展行动朴素分析理论。海德认为人有两种强烈的动机:一是形成对周围环境一贯性理解的需要;二是控制环境的需要。事件的原因无外乎有两种:一是内因,比如情绪、态度、人格、能力等;二是外因,比如外界压力、天气、情境等。一般人在解释别人的行为时,倾向于性格归因;在解释自己的行为时,倾向于情景归因。

海德还指出,在归因的时候,人们经常使用两个原则:一是共变原则(principle of covariation),它是指某个特定的原因在许多不同的情境下和某个特定结果相联系,该原

因不存在时,结果也不出现,我们就可以把结果归于该原因,这就是共变原则。二是排除原则,它是指如果内外因某一方面的原因足以解释事件,我们就可以排除另一方面的归因。

归因理论的指导原则和基本假设是:寻求理解是行为的基本动因。常见的归因理论还有韦纳的归因理论、阿布拉姆森等的归因理论、凯利的归因理论、琼斯和戴维斯的归因理论。

2.6 内在动机

1960年,布鲁纳(J. S. Bruner)在《教育过程》一书中强调了“内部动机”的作用,认为内在动机是推动学习的真正动力^[19],自此,人们开始重视内部动机对学习的影响。美国社会心理学家阿玛布丽(T. M. Amabile)的大量研究证明^[20],内部动机对人的创造性具有很大的促进作用。高水平的内在动机是杰出的创造性人才的重要特征。内在动机导致科学家把专业研究当成自己的事业,制定为之奋斗的自我目标。人们认为“认知好奇心”是内在动机的核心,这是一种追求外界信息、指向学习活动本身的内驱力,它表现为好奇、探索、操作和掌握行为。人们也把它称之为认知动机。

大多数内在动机的计算模型采用面向任务和基于经验。辛格(S. Singh)等人设计的内在动机强化学习智能体^[21],可以开发各种能力的应用。在这个模型中,智能体通过编程,以确定明显有趣事件的光线变化和声音强度。内在动机强化学习模型在多任务学习中很重要,重点不在如何定义动机。与此相反,心理学的内在动机理论提供了一个简洁的、与领域和任务无关的理论,其基础是动机的计算模型。

3 动机学习

3.1 觉知

觉知开始于外界刺激的输入,激活感知系统的初级特征检测器。输出信号被发送到感觉记忆中,在那里更高层次的功能探测器用于更抽象的实体,如对象、类别、行动、事件等的检测。所产生的知觉移动到工作区,在那里产生本地联系的短暂情景记忆和陈述性记忆会被做线索标记。这些本地联系与知觉结合,产生当前情景模型,用以表示智能体对当前正在发生的事情的理解。

在心智模型CAM中,觉知基本上是从感测到状态的感觉组合。智能体在复杂的环境中有效地工作,必须选择这些组合的一个子集作为觉知值。觉知函数是所检测到的感知状态 $S_{(t)}$ 映射到觉知的一个子集。

定义1(觉知函数) 觉知函数定义为由进一步处理的感觉状态的感知组合,包含较少的将影响智能体关注点的感觉信息,限制其状态空间所关注的子集。其中,典型的觉知函数 $A_{S_{(t)}}$ 可以表示为

$$A_{S_{(t)}} = \{(a_{1(t)}, a_{2(t)}, \dots, a_{L(t)}, \dots) \mid a_{L(t)} = s_{L(t)} (\forall L)\} \quad (1)$$

这意味着每个觉知在时间 t 关注感觉到的状态每个元素。 L 是感觉到的元素的长度,它是可变的。

3.2 事件

引入事件来建模觉知状态之间的转换。事件被表示为两种感觉状态之间的差值。感觉到的两种状态, $S_{(t')} = (s_{1(t')}, s_{2(t')}, \dots, s_{L(t')}, \dots)$ 和 $S_{(t)} = (s_{1(t)}, s_{2(t)}, \dots, s_{L(t)}, \dots)$, 其中 $t' < t$, 使用差异函数 Δ 计算状态之间的相差, 差异函数 Δ 定义如下。

定义 2(差异函数) 差异函数是分配一个值表示两种感觉 $S_{L(t)}$ 和 $S_{L(t')}$ 之间的差异, 在感测的状态 $S_{(t)}$ 和 $S_{(t')}$ 如下:

$$\Delta(S_{L(t)}, S_{L(t')}) = \begin{cases} S_{L(t)}, & \neg \exists S_{L(t')} \\ S_{L(t')}, & \neg \exists S_{L(t)} \\ S_{L(t)} - S_{L(t')}, & S_{L(t)} - S_{L(t')} \neq 0 \\ 0, & \text{否则} \end{cases} \quad (2)$$

差异函数提供的信息反映相继感知状态之间的变化大小。

定义 3(事件函数) 事件函数定义为智能体识别事件的差异变量的组合, 每个事件只包含一个非零的差异变量。事件函数可以被定义为如下公式:

$$E_{S(t)} = \{E_{L(t)} = (e_{1(t)}, e_{2(t)}, \dots, e_{L(t)}, \dots) \mid e_{e(t)}\} \quad (3)$$

其中

$$e_{e(t)} = \begin{cases} \Delta(S_{e(t)}, S_{e(t')}), & e = L \\ 0, & \text{否则} \end{cases} \quad (4)$$

事件可以是不同长度的甚至是空的, 这取决于感觉数量的变化。

3.3 新奇

检测新事件是任何信号分类方法的一个重要的功能。因为我们不能对机器学习系统训练所有可能遇到的对象类的的数据, 它就变得很重要, 它在测试时能够区分已知和未知的对象信息。新奇检测是一个非常具有挑战性的任务, 可以在复杂、动态的环境中发现新颖、感兴趣的事件。新奇检测是一个很好的分类或识别系统的基本要求, 因为有时候测试数据中包含的对象, 训练模型时信息并不知道。觉知的新颖性是关系到认知, 而认知的新颖性是关系到知识。基于固定组训练样本从一个固定数量的类别, 新奇的检测是一个二元决策任务对每个测试样本确定它是否属于一个已知的类别。

定义 4(新奇检测函数) 新奇检测函数 N , 采用智能体的概念状态, $c \in C$, 并与以前的经历记忆比较, $m \in M$, 通过长时记忆的建构产生一个新奇的状态 $n \in N$:

$$N: C \times M \rightarrow N \quad (5)$$

在 CAM 中, 采用由科霍南(T. Kohonen)提出的自组织映射神经网络(SOM)实现新奇检测, 这种网络是非监督的、竞争学习的聚类网络^[31]。科霍南认为, 神经网络在接受外界输入时, 将会分成不同的区域, 不同的区域对不同的模式具有不同的响应特征, 即不同的神经元以最佳方式响应不同性质的信号激励, 从而形成一种拓扑意义上的有序图。这种有序图也称为特征图, 它实际上是一种非线性映射关系, 将信号空间中各模式的拓扑关系几乎不变地反映在这张图上, 即各神经元的输出响应上。由于这种映射是通过无监督的自适应过程完成的, 所以也称它为自组织特征图。

3.4 兴趣度

兴趣度定义为新奇和惊喜,这取决于觉知当前的知识和计算能力。兴趣度可以是客观的或主观的:客观兴趣度使用关系完全在对象被认为是有趣的,而主观兴趣度比较对象的属性与用户确定利益的信念。一种情景的兴趣度是衡量对于智能体的现有知识情况的重要性,有趣的是这与先前的经验不太相似,或者大不一样。

定义 5(兴趣度函数) 兴趣度函数决定了情景的兴趣性价值, $i \in I$, 基于新奇检测, $n \in N$:

$$I: N \rightarrow I \quad (6)$$

注意是选择性地集中在环境的某个方面而忽视其他事情的行为和认知过程。根据兴趣度,采用阈值选择机制(TSM)^[22]。TSM 是一个阈值滤波算法。假设我们得到一个阈值, T , 如果兴趣度值大于 T 事件选择建立一个激励,引起注意;相反地,如果该值小于 T , 事件被省略。

3.5 注意

定义 6(注意选择) 注意是复杂的认知功能,这是人类行为的本质。注意是一个外部选择过程(声音,图像,气味……)或内部(思维)事件都必须保持在一定水平的觉知。根据给定的语境情况下,选择性或集中注意力的选择在信息上应优先处理。选择性注意使你专注于一个项目,而明智地识别和区分不相关信息。CAM 采用兴趣度策略来实现注意选择。

上面我们讨论的是觉知阶段,或对外界语境理解的注意,这对产生意识发挥重要作用。在意识形成后,另一类型的注意起作用,唤醒和协调脑区各部分功能,协同完成任务,达到期望的目标。

3.6 动机学习算法

动机学习是对观察到的感觉输入创建内部表示,并将它们链接到对其操作有用的学习动作。如果观察到的事件与当前目标不相关,没有兴趣,则不会引起注意,不发生动机学习。这种对学习内容的筛选是非常有用的,因为它节省机器的存储器,免于存储不重要的观察。基于新颖性的动机学习只是关注具有足够兴趣的事件。下面描述 CAM 中基于新颖性的动机学习算法。

算法 1 Motivation learning algorithm

- (1) Observe $O_{S(t)}$ from $S_{(t)}$ using the observation function
 - (2) Subtract $S_{(t)} - S_{(t')}$ using the difference function
 - (3) Compose $E_{S(t)}$ using the event function
 - (4) Look for $N_{(t)}$ using introspective search
 - (5) Repeat (for each $N_i(t) \in N(t)$)
 - (6) Repeat (for each $I_j(t) \in I(t)$)
 - (7) $Attention = \max I_j(t)$
 - (8) Create a Motivation by Attention.
-

4 基于动机的强化学习

目前,存在许多的基于动机的强化学习(motivated reinforcement learning, MRL)模型,以不同的方式对强化学习进行扩展。可以将 MRL 模型分为两大类:①将动机信号引入作为奖励信号的补充;②将奖励信号替换为动机信号^[4]。在这两类中,MRL 模型还可以按照它们整合的强化学习(reinforcement learning, RL)算法的类别来进一步划分。尽管将动机与函数逼近等其他算法结合也是可行的,但现有的工作主要集中在将动机整合进强化学习和层次强化学习(hierarchical reinforcement learning, HRL)中。另外,这些整合后的模型的设计目标也发生了变化。一些模型旨在加快 RL 和 HRL 算法的速度;其余模型则通过将动机作为一种自动注意专注机制来实现更具自适应性和多任务的学习。

这里,术语“动机信号”被用来区分奖励信号,该信号使用动机计算模型作为一种基于智能体经验的函数进行在线计算,而不是用来作为一组预定义的规则映射已知环境状态或状态转换间的值。如图 2 所示的 I 型模型,将动机信号 $R_{m(t)}$ 整合进来作为奖励信号 $R_{(t)}$ 的补充,并专注于动机在 RL 和 HRL 中的应用。这些模型的设计目标除了通过加入动机信号来加快现有 RL 算法的速度,还包括通过将动机作为一种自动注意专注机制来实现更具自适应性和多任务的学习。动机层次强化学习(motivated hierarchical reinforcement Learning, MHRL)模型将识别子任务的过程形式化为一种动机过程,该过程产生一个动机信号 $R_{m(t)}$ 作为来自环境的标准奖励信号 $R_{(t)}$ 的补充。动机信号的首要目标是通过识别由奖励信号所定义任务的子任务来指导学习。I 型模型严格地遵守了 RL 和 HRL 的目标,即加快奖励任务的学习速度。

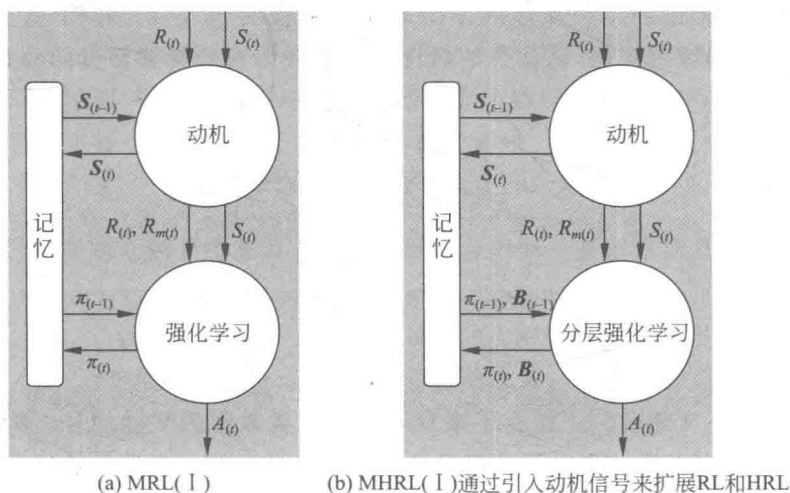


图 2 I 型 MRL 模型

如图 3 所示,II 型 MRL 模型直接将奖励信号 $R_{(t)}$ 替换为动机信号 $R_{m(t)}$ 。这种方法主要用在与 RL 算法的结合中。尽管有大量的 HRL 和 MHRL(I)模型声称能扩展到