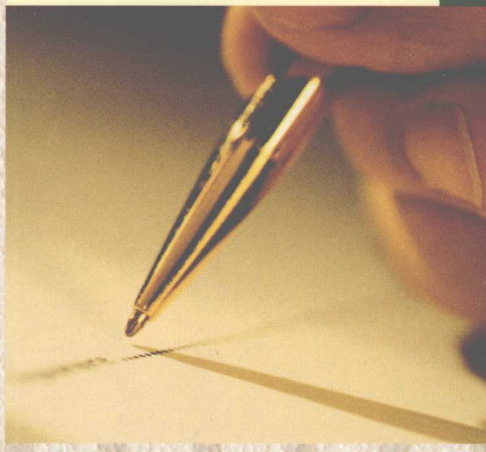
学者书屋系列

分层强化学习理论与方法

沈 晶◎编著



哈尔滨工程大学基础研究基金资助项目

Harbin Engineering University Press
哈尔滨工程大学出版社

哈尔滨工程大学基础研究基金资助项目

学者书屋系列

分层强化学习理论与方法

沈　晶　编著

顾国昌　主审

哈尔滨工程大学出版社

内容简介

强化学习通过试错与环境交互获得策略的改进，其自学习和在线学习的特点使其成为机器学习研究的一个重要分支。但是，强化学习一直被维数灾难所困扰，近年来，分层强化学习在克服维数灾难方面取得了显著进展。本书系统地介绍了强化学习、分层强化学习的理论基础和学习算法以及作者在分层强化学习领域的研究成果和该领域的最新研究进展。

本书可作为高等院校和科研机构从事计算机应用、人工智能和机器学习等相关专业和方向的教师、研究人员、研究生及高年级本科生参考使用。

图书在版编目(CIP)数据

分层强化学习理论与方法/沈晶编著. —哈尔滨：哈尔滨工程大学出版社，2007.12

ISBN 978-7-81133-028-1

Ⅰ.分… Ⅱ.沈… Ⅲ.机器学习-研究 Ⅳ.TP181

中国版本图书馆 CIP 数据核字(2007)第 194194 号

出版发行 哈尔滨工程大学出版社
社　　址 哈尔滨市南岗区东大直街 124 号
邮政编码 150001
发行电话 0451-82519328
传　　真 0451-82519699
经　　销 新华书店
印　　刷 哈尔滨工业大学印刷厂
开　　本 787mm×960mm 1/16
印　　张 9.5
字　　数 110 千字
版　　次 2007 年 12 月第 1 版
印　　次 2007 年 12 月第 1 次印刷
定　　价 19.00 元

http://press.hrbeu.edu.cn
E-mail:heupress@hrbeu.edu.cn

前　言

强化学习通过试错与环境交互获得策略的改进，其自学习和在线学习的特点使其成为机器学习研究的一个重要分支。但是，强化学习一直被维数灾难所困扰。近年来，分层强化学习在克服维数灾难方面取得了显著进展，典型的成果有 Option，HAM 和 MAXQ 等方法，其中 Option 和 MAXQ 方法在目前使用较为广泛。Option 方法便于自动划分子任务（尤其分区或分段子任务），且子任务粒度易于控制，但利用先验知识划分子任务时，任务划分结果表达不够明晰，且子任务内部策略难于确定；MAXQ 方法在线学习能力强，但自动分层能力较弱，且分层粒度不够精细，难以对一些规模很大的子任务作出进一步的分解。本书在系统地介绍了强化学习、分层强化学习的理论基础和学习算法之后，探讨了一种集成 Option 和 MAXQ 的分层强化学习新方法——OMQ，并深入研究集成过程中所涉及的理论与计算问题，以及该方法在动态环境、多智能体环境中应用时需要进一步解决的问题。

本书得到了总装备部预研基金及哈尔滨工程大学基础研究基金（HEUFT07022，HEUFT05021，HEUFT05068）的资助，在编写过程中，得到了哈尔滨工程大学计算机科学与技术学院顾国昌教授和张国印教授的悉心指导，以及刘海波博士的鼎力相助，张汝波教授审阅了本书初稿，提出了宝贵的意见，哈尔滨工程大学出版社的编辑老师付出了艰辛的劳动，在此一并表示感谢！

由于作者的水平有限，加之时间仓促，书中难免存在一些不足和错误之处，欢迎读者批评指正。

沈　晶

2007 年 12 月

目　录

第1章 绪 论

分层强化学习是在强化学习的基础上通过增加“抽象机制”而形成的一种效率更高的机器学习方法。本章将对机器学习、强化学习、分层强化学习的定义以及研究现状进行介绍。

1.1 机器学习

1.1.1 机器学习的定义

机器学习(Machine Learning)的核心是学习。学习是人类具有的一种重要智能行为,但究竟什么是学习,长期以来却众说纷纭。这是因为进行这一研究的人们分别来自不同的学科,更重要的是学习是一种多侧面、综合性的心理活动,它与记忆、思维、知觉、感觉等多种心理行为都有着密切的联系,人们难以把握学习的机理与实现。社会学家、逻辑学家和心理学家都各有其不同的看法。按照人工智能大师Simon的观点,学习就是系统在不断重复的工作中对本身能力的增强或者改进,使得系统在下一次执行同样任务或类似任务时,会比现在做得更好或效率更高。这一阐述包含过程、系统与改进性能这样三个要点。学习的基本模型就是基于这一观点建立起来的。

机器学习至今还没有统一的定义,而且也很难得到一个公认的和准确的定义。顾名思义,机器学习是研究如何使用机器来模拟人类学习活动的一门学科。稍微严格的提法:机器学习是一门研究机器获取新知识和新技能,并识别现有知识的学问。

从计算机科学的角度来看,有的学者认为可以将机器学习分为工程的观点和科学的观点两种。工程的观点认为机器学习是人类学习的计算机实现;科学的观点认为机器学习是人类学习的计算机模拟。这两种观点的区别主要在于:实现是指完成相同的功能,模拟是指把握相同的原理。目前人类对自身的思维规律和学习奥秘仍然知道甚少,所以,要达到人类学习的计算机模拟还不太现实。目前,人工智能界的主流观点倾向于工程的观点。计算机科学家们主要关注不同系统实现机器学习的过程以及性能改进的效果。

机器学习是使计算机具有智能的根本途径。正如 Shank 所说:"一台计算机如果不会学习,就不能称之为具有智能。"

机器学习是机器具有智能的重要标志,同时也是机器获得知识的根本途径,所以,机器学习在机器智能中占有重要地位。自 20 世纪 80 年代以来,机器学习作为继专家系统之后人工智能应用的又一重要的研究领域,在人工智能界引起了广泛的关注,它已成为人工智能界的重要课题之一。

机器学习的目标及研究工作主要包括如下几个方面:

(1)面向任务的研究:研究和分析改进一组预定任务的执行性能的学习系统;

(2)认知模拟:研究人类学习过程并进行计算机模拟,由此建造高性能的计算机系统;

(3)理论性分析:从理论上探索各种可能的学习方法和独立于应用领域的算法,加强机器学习的理论背景研究,规范机器学习的技术、方法和理论。

1.1.2 机器学习的发展史

机器学习是人工智能研究领域较为年轻的分支。回顾它的发展历程,可以将其划分为不同的阶段。如何划分这些阶段,有不同的方

法。例如,可以按照机器学习的研究途径和目标,将其划分为神经元模型研究、符号概念获取、知识强化学习、连接学习和混合型学习五个阶段;也可以按照机器学习的发展过程,将其划分为热烈时期、冷静时期、复兴时期和蓬勃发展时期。下面按照后一种划分方法对其各个阶段分别介绍。

第一阶段是从20世纪50年代中叶至20世纪60年代中叶,属于热烈时期。在这个时期,所研究的是"没有知识"的学习,即"无知"学习,它的主要研究目标是研制各类自组织和自适应系统。例如,如果给系统一组刺激、一个反馈源、以及修改它们自身组织的足够自由度,那么它们将改变自身成为最优的组织,即它们能够修改自身以适应它们的环境。这类系统主要采用的研究方法是不断修改系统的控制参数以改进系统的执行能力,不涉及面向具体问题的知识。

这一阶段研究的理论基础是20世纪40年代就有的神经网络模型。有的学者将机器学习的起点定为1943年,即McCulloch与Pitts对神经元模型(简称为MP模型)的研究[1]。他们研究的历史意义是在科学发展进程中,首次发现了人类神经元的工作方式,并给出了这种工作方式的数学描述。这项研究在科学史上的意义是非同寻常的,它第一次揭示了人类神经系统的工作方式;这项研究对近代信息技术发展的影响也是巨大的,计算机科学与控制理论均从这项研究中受到了启发。由于Pitts的努力,使得这项研究的结论没有仅仅停留在生物学的领域内,他为神经元的工作方式建立了数学模型,正是这个数学模型深刻地影响了机器学习的研究。电子计算机的产生和发展,使得机器学习的实现成为可能。人们研制了各种模拟神经的计算机,其中Rosenblatt的感知机最为著名,它由阈值性神经元组成,试图模拟人脑的感知及学习能力[2]。遗憾的是,大多数希望产生某些复杂智能系统的企图都失败了。不过,这一阶段的研究导致了模式识别这门新学科的诞生,同时,形成了机器学习的两种重要方法,

即判别函数法和进化学习法。著名的 Samuel 下棋程序就是使用判别函数法的典型代表。该程序具有一定的自学习、自组织、自适应能力，能够根据下棋时的实际情况决定走步策略，并且从经验中学习，不断地调整棋盘局势评估函数，在不断的对弈中提高自己的棋艺。四年后，这个程序战胜了设计者本人。又过了三年，这个程序战胜了美国一个保持 8 年之久的常胜不败的冠军。不过，这种脱离知识的感知型学习系统具有很大的局限性。无论是神经模型、进化学习还是判别函数法，所取得的学习结果都是很有限的，它们远远不能满足人类对机器学习系统的期望。在这一阶段，我国研制了数字识别学习机。

第二阶段是从 20 世纪 60 年代中叶至 20 世纪 70 年代中叶，被称为机器学习的冷静时期。这一阶段的主要研究目标是模拟人类的概念学习过程。机器的内部表示采用逻辑结构或图结构，机器采用符号来表示概念，又称符号概念获取，并提出关于所学概念的各种假设。在此阶段，研究者意识到学习是复杂而困难的过程，因此，人们不能期望学习系统可以从没有任何知识的环境中开始，学习到高深而有价值的概念。这种观点使得研究人员一方面深入探讨简单的学习问题，另一方面则把大量的领域专家知识加入到学习系统中。

这一阶段具有代表性的工作有 Winston 的结构学习系统和 Hayes-Roth 等人的基于逻辑的归纳学习系统。1970 年，Winston 建立了一个从例子中进行概念学习的系统，它可以学会积木世界中一系列概念的结构描述[3]。尽管这类学习系统取得了较大的成功，但是所学到的概念都是单一概念，并且大部分都处于理论研究和建立实验模型阶段。除此之外，神经网络学习机因理论缺陷未能达到预期效果而转入低潮。因此，那些曾经对机器学习的发展抱有极大希望的人们对此感到很失望。人们又称这个时期为机器学习的“黑暗时期”。

第三阶段是从20世纪70年代中叶至20世纪80年代中叶,称为复兴时期。这一阶段的主要研究目标仍然是模拟人类的概念学习过程,但是研究者已经从学习单个概念扩展到学习多个概念,探索不同的学习策略和各种学习方法。

机器的学习过程一般都是以大规模的知识库作为背景,实现知识强化学习。值得高兴的是,这一阶段研究者开始将学习系统与各种应用系统结合起来,并获得了极大的成功,在实际应用中发挥了重要作用。同时,专家系统在知识获取方面的需求,也极大地刺激了机器学习的研究和发展。在出现第一个专家学习系统之后,示例归纳学习系统成为研究的主流,自动知识获取成为机器学习应用的研究目标。

1980年,在美国的卡内基·梅隆召开了第一届机器学习国际研讨会,标志着机器学习研究已在全世界兴起。此后,机器学习开始得到了大量地应用。Strategic Analysis 和 Information System 国际杂志连续三期刊登有关机器学习的文章。1984年,由 Simon 等20多位人工智能专家共同撰文编写的 Machine Learning 文集第二卷出版,国际性杂志 Machine Learning 的创刊,这些事件更加显示出机器学习突飞猛进的发展趋势。这一阶段代表性的工作有 Mostow 的指导式学习,Lenat 的数学概念发现程序 AM,Langley 的 BACON 程序及其改进程序,它们可以根据经验领域的原始数据发现一些基本的物理学定律和化学定律。其他比较著名的归纳学习方法有 Quinlan 的 ID3 算法,Michalski 的星算法及其概念聚类思想;在基于解释学习系统中,有 DeJong 的 Genesis 系统,Mitchell 的 LEX 系统和 Minton 的 Prodigy 系统等;在类比学习中,Winston,Carbonell 和 Gentner 等人也做了许多开拓性的工作[4]。

这一阶段的研究特点主要包括以下三点:

(1)基于知识的方法,即首先具备大量初始知识;

(2)开发出各种各样的学习方法,包括示教学习、观察和发现学习、类比学习以及解释学习等;

(3)结合生成和选择学习任务的能力,应用了启发式信息。

第四阶段始于1986年,是其蓬勃发展的时期。这一阶段是机器学习发展的最新阶段。一方面,神经网络的研究重新兴起。在此前的10多年中,虽然神经元模型研究陷入低潮,但仍有一部分学者在潜心研究。他们不懈地努力,终于克服了神经元模型的局限性,提出了多层网络的学习算法,再加上VLSI技术、超导技术、生物技术、光学技术等的发展与支持,神经元网络研究又在一个新的起点上再度兴起,从而使机器学习进入连接学习的阶段。目前对连接学习方法的研究方兴未艾,机器学习的研究已在全世界范围内出现新的高潮,关于机器学习的基本理论和综合系统的研究得到了加强和发展。另一方面,实验研究和应用研究受到前所未有的重视。随着计算机技术和人工智能技术的迅猛发展,机器学习有了新的更有力、更有效的研究手段和研究环境。例如,这一阶段的符号学习由"无知"学习转向有专门领域知识的增长型学习,因而,出现了具有一定领域知识的分析学习。

在连接学习重新兴起的同时,传统的符号学习研究也取得了很大的进展。实际上,连接学习和符号学习各有所长,并具有很大的互补性。因此,把符号学习和连接学习结合起来的混合型学习系统研究已成为一个新的热点,如果能够把这两种不同的学习机制有机地融合在一起,就可以在一定程度上有机地模拟人类的逻辑思维和直觉思维,这将是人工智能领域的一个重大突破。目前,研究者已经提出了一些混合方法,这些方法的基本思路是将符号学习所学到的不完善的领域知识按照一定的转化规则构成一个神经网络,然后再利用连接机制继续学习。

从国内外的研究现状来看,将上述两种学习机制结合,无论是理

论研究还是实际应用都有着广阔的发展前景。例如,基于生物发育进化论的进化学习系统和遗传算法,吸收了归纳学习与连接机制学习的长处而受到重视。基于行为主义的强化学习系统因吸收了连接机制和遗传算法的思想而显示出了新的生命力。这一阶段代表性的工作有 Rulmelhant 的 BP 模型,Hopfield 的 Hopfield 模型,Kohnen 的 Kohnen 模型,Holland 的遗传算法,Newell 的经验学习系统 SOAR,Michalski 的示例学习系统 AQl5,Watkins 的 Q - 算法,Sutton 的强化学习 TD 算法,Vapnik 的统计机器学习,Lamma 等的多策略学习方法等[4,5]。

这一阶段机器学习具有以下显著特点:

(1)机器学习已成为新的边缘学科,许多高校已将机器学习作为一门课程。它综合应用了心理学、生物学和神经生理学以及数学、自动化和计算机科学,形成了机器学习理论基础;

(2)开发出了各种各样的学习方法,各种学习方法的应用范围不断扩大,相当一部分已成为了商品。归纳学习的知识获取工具已在诊断专家系统中得到广泛使用。连接学习在语音识别和图像识别中占有优势。分析学习已用于设计综合型专家系统。遗传算法与强化学习在工程控制中有着较好的应用。与符号系统耦合的神经网络连接学习在企业的智能管理与智能机器人运动规划中发挥作用;

(3)结合各种学习方法,将多种学习方法综合集成的系统研究正在兴起。尤其是连接学习和符号学习的耦合,可以更好地解决连续性信号处理中知识和技能的获取与求精问题,因而受到重视;

(4)机器学习与人工智能各种基础问题的统一性观点正在形成;

(5)与机器学习的有关学术活动空前活跃。世界上每年都要召开机器学习的研讨会,还有计算学习理论会议、神经网络大会以及遗传算法会议。近十多年来,我国的机器学习研究开始稳步发展和逐步繁荣。每两年举办一次全国性的机器学习研讨会,学术讨论和科

技开发蔚然成风。

目前,机器学习的研究已不仅是人工智能领域的重要问题,而且已经成为计算机科学的核心问题,并提出了如下几个迫切需要解决的问题。

(1)计算的个性化,即对个人需求适应的计算。它涉及许多复杂的问题,它必须解决计算机对用户需求的适应计算问题。这种适应性计算是建立在指令空间还是建立在情感空间将产生两类完全不同的计算系统。

(2)由于机器学习中的许多算法受启发于认知心理学与神经生理学等非精确科学,这些算法或多或少地存在随意性,理论描述较为缺乏。使用更精确的数学方法深入地研究机器学习中的理论问题,已是当务之急。

(3)对结构化和非结构化海量数据的理解,即所发展的机器学习算法必须能够解决海量数据的理解问题,这是开展机器学习研究和评价研究结果的重要条件。

机器学习已经成为一门新的边缘学科,它与认知科学、神经心理学、逻辑学、教育学和哲学等学科都有着密切的联系,并对人工智能的其他分支,例如:专家系统、自然语言理解、自动推理、智能机器人、计算机视觉、计算机听觉等方面,都起到了重要的推动作用。因此,机器学习必将具有十分广泛的应用前景。

1.1.3 机器学习系统的基本模型

Simon 对学习的阐述只是对机器学习的一个一般性的概述,只是一种理念。根据 Simon 对学习的阐述,一个学习系统应该满足如下基本的要求。

(1)具有合适的学习环境

这里所说的学习环境,是指学习系统进行学习时的信息来源。

如果把学习系统类比为学生的话,那么学习环境则是为学生提供信息的教师、书本以及各种实验条件等。毫无疑问,没有学习环境,学生就不可能学习到新的知识以及运用所学到的知识来解决问题。

(2)具有一定的学习能力

学习环境为学习系统提供了相应的信息和基础,学习系统还必须具备一定的学习能力和适当的学习方法,否则也学不到知识或者不会有好的学习效果。正如在同样的学习环境中,不同的学生,他们的学习能力和学习方法不同,他们的学习效果也往往大相径庭。

(3)能够运用所学到的知识来求解问题

学习系统之所以有意义有价值,是在于可以学以致用。和人类学习一样,一个学习系统如果不能将所学到的知识用于实际问题的解决,那么学习也就失去了其最重要的作用和意义。学习系统应该能够将所学到的信息用于未来的估计、分类、决策和控制,以便改进系统的性能。

(4)通过学习提高自身的性能

在 Simon 的阐述中,改进系统性能是学习的三个要点之一。一个学习系统应该能够通过学习增长知识,提高技能,改进性能,使自己能够做一些原来无法做到的事,或者可以将原先能够做到的事做得更好。

通过以上分析,可以得出一个学习系统至少应该包括这样四个重要环节:环境、学习单元、知识库和执行单元,它们之间的关系如图 1-1 所示。在具体应用中,环境、知识库和执行单元决定了具体的工作内容,学习单元所需要解决的问题就由这三部分确定。

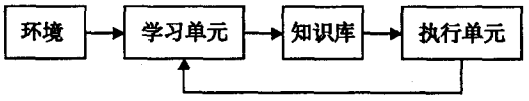


图 1-1 机器学习的基本模型