



中国计算机学会  
学术著作丛书

# 基于解释的机器学习方法

石纯一等 著

清华大学出版社  
广西科学技术出版社

中国计算机学会学术著作丛书

# 基于解释的机器学习方法

石纯一等著

清华大学出版社  
广西科学技术出版社

(京)新登字 158 号

### 内 容 简 介

基于解释的机器学习是一种分析式学习方法,是人工智能学科的一个重要分支。它的特点是依知识库和少量例子通过演绎过程进行学习。

本书介绍基于解释机器学习的研究现状和基本方法。全书共 7 章,依次介绍基于解释的机器学习模型、可操作性、领域理论和效用分析等内容。

本书可作为大学本科生、研究生学习人工智能课程的参考书,也可供从事计算机科学、人工智能等方面的科技人员参考。

版权所有,翻印必究。

本书封面贴有清华大学出版社激光防伪标签,无标签者不得销售。

书 名: 基于解释的机器学习方法

作 者: 石纯一 等

出版者: 清华大学出版社(北京清华大学校内,邮编 100084)

广西科学技术出版社(南宁市河堤路 14 号,邮编 530021)

印刷者: 人民文学印刷厂

发行者: 新华书店总店北京科技发行所

开 本: 787×1092 1/16 印张: 11.25 字数: 267 千字

版 次: 1997 年 4 月第 1 版 1997 年 4 月第 1 次印刷

书 号: ISBN 7-302-02432-4/TP·1230

印 数: 0001—3000

定 价: 16.00 元

## 序 言

计算机是当代发展最为迅猛的科学技术,其应用几乎已深入到人类社会活动和生活的一切领域,大大提高了社会生产力,引起了经济结构、社会结构和生活方式的深刻变化和变革,是最为活跃的生产力之一。计算机本身在国际范围内已成为年产值达 2500 亿美元的巨大产业,国际竞争异常剧烈,预计到本世纪末将发展为世界第一大产业。计算机科技具有极大的综合性质,与众多科学技术相交叉而反过来又渗入更多的科学技术,促进它们的发展。计算机科技内容十分丰富,学科分支生长尤为迅速,日新月异,层出不穷。因此在我国计算机科技尚比较落后的情况下,加强计算机科技的传播实为当务之急。

中国计算机学会一直把出版图书刊物作为学术活动的重要内容之一。我国计算机专家学者通过科学实践,做出了大量成果,积累了丰富经验与学识。他们有撰写著作的很大积极性,但相当时期以来计算机学术著作由于印数不多,出版往往遇到不少困难,专业性越强越有深度的著作,出版难度越大。最近清华大学出版社与广西科学技术出版社为促进我国计算机科学技术及产业的发展,推动计算机科技著作的出版工作,特设立“计算机学术著作出版基金”,以支持我国计算机科技工作者撰写高水平的学术著作,并将资助出版的著作列为中国计算机学会的学术著作丛书。我们十分重视这件事,并已把它列为学会本届理事会的工作要点之一。我们希望这一系列丛书能对传播学术成果、交流学术思想、促进科技转化为生产力起到良好作用,能对我国计算机科技发展具有有益的导向意义,也希望我国广大学会会员和计算机科技工作者,包括海外工作和学习的神州学人们能积极投稿,出好这一系列丛书。

中国计算机学会

1992 年 4 月 20 日

## 出版说明

近年来,随着微电子和计算机技术渗透到各个技术领域,人类正在步入一个技术迅猛发展的新时期。这个新时期的主要标志是计算机和信息处理的广泛应用。计算机在改造传统产业,实现管理自动化,促进新兴产业的发展等方面都起着重要作用,它在现代化建设中的战略地位愈来愈明显。计算机科学与其它学科的交叉又产生了许多新学科,推动着科学技术向更广阔的领域发展,正在对人类社会产生深远的影响。

科学技术是第一生产力。计算机科学技术是我国高科技领域的一个重要方面。为了推动我国计算机科学及产生的发展,促进学术交流,使科研成果尽快转化为生产力,清华大学出版社与广西科学技术出版社联合设立了“计算机学术著作基金”,旨在支持和鼓励科技人员,撰写高水平的学术著作,以反映和推广我国在这一领域的最新成果。

计算机学术著作出版基金资助出版的著作范围包括:有重要理论价值或重要应用价值的学术专著;计算机学科前沿探索的论著;推动计算机技术及产业发展的专著;与计算机有关的交叉学科的论著;有较大应用价值的工具书;世界名著的优秀翻译作品。凡经作者本人申请,计算机学术著作出版基金评审委员会评审通过的著作,将由该基金资助出版,出版社将努力做好出版工作。

基金还支持两社列选的国家高科技重点图书和国家教委重点图书规划中计算机学科领域的学术著作的出版。为了做好选题工作,出版社特邀请中国计算机学会、中国中文信息学会帮助做好组织有关学术著作丛书的列选工作。

热诚希望得到广大计算机界同仁的支持和帮助。

清华大学出版社  
广西科学技术出版社 计算机学术著作出版基金办公室

1992年4月

清华大学出版社 广西科学技术出版社  
计算机学术著作出版基金

## 评审委员会

主任委员	张效祥				
副主任委员	汪成为	唐泽圣			
委 员	王鼎兴	杨芙清	李三立	施伯乐	徐家福
	夏培肃	董韫美	黄 健	焦金生	

## 前 言

基于解释的机器学习(Explanation-Based Learning, 简称 EBL)是一种分析式学习方法,特点是从少数例子出发,依领域理论(知识库)通过演绎获得知识。自 20 世纪 80 年代中 EBL 形成一种机器学习方法以来,国内外已开展了一系列的研究,并已有初步应用。

清华大学计算机系机器学习课题组,自 1985 年以来已进行了 10 年的 EBL 研究,先后完成了国家自然科学基金 2 项,国家高技术 863-306 基础性研究课题 2 项,共发表论文 20 余篇,所做研究成果经鉴定达到国际先进水平。

这本书涉及解释学习的主要内容,包括 EBL 的可操作性、领域理论和效用问题等。有些内容除在我们的讨论班上报告过外,还在国家智能研究开发中心报告过,也得到不少国内人工智能学术界年轻学者的帮助。这本书是集体研究成果的汇集,是我们研究工作的总结,包括张旗博士,郝继刚博士,王建伟博士,以及陈红、龚义涛、郭晨东、黄毅青、杨宇、于津、王彤等硕士的研究工作。

本书最后由郝继刚、王建伟整理完成。在课题组的讨论班上得到了陆玉昌教授、李膺春教授的指导,周远晖、张朝辉博士生也参加了讨论,一并表示感谢。

在我国计算机界的著名学者张效祥先生、董韫美先生、陆汝钫先生的支持下,本书获得了清华大学出版社、广西科技出版社的计算机学术著作出版基金的资助,在此也向这几位先生和出版社深表谢意。

书中一些观点和见解定会有不少不准确甚至不当之处,请读者批评指正。

石纯一

1996 年 8 月于清华园

# 目 录

<b>第 1 章</b>	<b>解释学习概述</b>	1
1.1	机器学习	1
1.2	解释学习	10
1.3	我们的工作	28
<b>第 2 章</b>	<b>可操作性</b>	30
2.1	可操作性的提出	30
2.2	可操作性的研究	31
2.3	可操作性的讨论	36
2.4	EBL 可操作性的模糊集描述及计算	37
2.5	讨论	41
<b>第 3 章</b>	<b>数字概括</b>	42
3.1	数字概括的目标分解法	42
3.2	数字概括的效用分析	47
<b>第 4 章</b>	<b>领域理论</b>	54
4.1	不完善领域理论问题的分类与表示(一)	54
4.2	不完善领域理论问题的分类与表示(二)	61
4.3	不完善领域理论的知识求精	66
4.4	基于多例多概念的领域理论修正方法	73
4.5	克服领域理论不完全问题的学习方法 CE	77
4.6	多策略、交互式的知识求精系统 MS/I-KRS	86
<b>第 5 章</b>	<b>效用问题</b>	100
5.1	效用分析	100
5.2	效用提高策略	104
5.3	进一步研究的课题	122
<b>第 6 章</b>	<b>解释学习算法的扩充</b>	123
6.1	模糊概念的学习	123
6.2	概念层次中的多策略学习	128
6.3	增量式的解释学习	130
6.4	算法分析与改进	136
6.5	有噪声的学习算法	141
<b>第 7 章</b>	<b>专家判断规则的获取方法</b>	146



7.1 引言 .....	146
7.2 专家决策认知模型 MBH .....	146
7.3 Polynome 系统 .....	148
7.4 IAMBH 系统 .....	149
<b>参考文献</b> .....	<b>156</b>

# 第 1 章 解释学习概述

## 1.1 机器学习

### 1.1.1 机器学习的研究内容

#### 1.1.1.1 何谓机器学习

获得新认识和适应新环境是智能体的本质特征之一。同智能体一样,学习能力目前只能从人类和高等生物那里得到“存在性”验证,尚未形成一致的形式化的定义和理论。机器学习(Machine Learning, ML)是人工智能(AI)的基本问题之一。

关于学习的定义,主要有:

(1) 学习是系统内部的适应性变化,使系统在以后从事同一任务或同一问题范围中类似的任务时,效率更高[Simon,1983];

(2) 学习是我们心智内部有用的变化[Minsky,1985];

(3) 学习是对正在被经历的事物,构造或修改其表示[Michalski,1986];

(4) 经由学习活动而使系统获得的能力,能够完成学习之前无法完成的新任务,或将老任务做得更好[Carbonell,1990];

(5) 机器学习是计算机系统有目的地增进其性能的过程[卢美律等,1995]。

综上所述,可以归纳出关于学习活动的几个特征:

(1) 目的性:学习活动是一种有目的的变化,目标是改善智能体的问题求解能力和性能。

(2) 信息来源:学习活动的输入信息包括外部环境信息和系统从前的经验两大类。

(3) 记忆能力:学习活动的结果体现和保存于智能体中,不随学习活动的完成而消失。

(4) 表示形式:一般包括显式知识获得、知识重组与重构、系统结构变化、系统参数调整、运行机制变化等。

#### 1.1.1.2 机器学习的研究内容和理论模型

通常根据学习目的的不同,可以把机器学习系统分为两类:一类是综合性的,主要目标是获取新知识(Knowledge Acquisition);另一类是分析性的,目标是对知识进行重组和形式变换,使之效率更高,称为技能改进(Skill Refinement),如图 1.1 所示。

Simon 给出一个抽象的学习模型,有四个构成要素:环境、学习单元、知识库和执行单元,见图 1.2 [Simon,1983]。环境为学习单元提供某些信息,学习单元利用这些信息对知识库进行某种改进,执行单元利用这些信息去执行任务,其间获得的信息可以反馈给学习单元,也可以对环境做出响应。

20 世纪 90 年代初 Michalski 等人提出学习推理理论,用抽象的知识变形算子分析、解释不同的学习方法和学习系统。主要观点有:

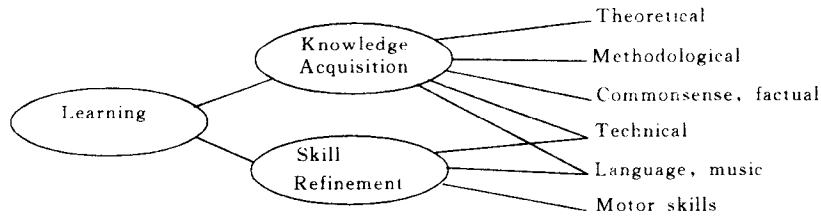


图 1.1 基本的学习形式和所学的知识或技能

- (1) 从推理中学习, 学习 = 推理 + 记忆。
- (2) 变形作为学习算子, 将学习过程定义为:  
 Input: 外界输入信息(I); 学习目标(G); 系统已具备的基础知识(BK); 变换(T)。  
 Output: 输出满足G的知识(O), 依BK将T作用于I来实现。
- (3) 对学习过程的不同分类标准中, 最关键的是学习目的。

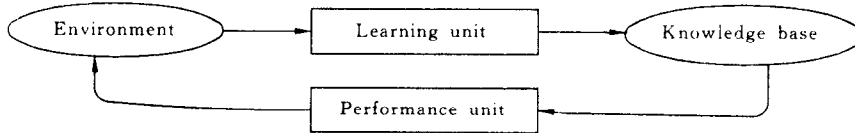


图 1.2 学习模型

### 1.1.2 机器学习的发展历史

机器学习的研究历史, 几乎与人工智能一样长, 大致经历了三个阶段:

第一个阶段, 是 20 世纪 50 年代的神经模拟和决策理论技术, 学习系统在运行时还很少具有结构或知识。主要方法是建造神经网络和自组织系统, 学习表现为阈值逻辑单元传送信号的反馈调整。著名系统有感知机(Perceptron) ([Rosenblatt, 1958]) 和西洋跳棋程序(checker) ([Cohen et al, 1982]) 等。随后的理论研究揭示, 这种不要知识的感知器类型的学习系统有很大的局限性。

第二阶段, 是 20 世纪 60 年代早期开始研究面向概念的学习, 即符号学习。使用的工具是语义网络或谓词逻辑, 不再是数值或统计方法。在概念获取中, 学习系统通过分析相关概念的大量正例和反例来构造概念的符号表示。在这一阶段, 人们认识到学习是个复杂而循序渐进的过程; 如果不要任何初始知识, 则学习系统无法学到高层次的概念。关于积木世界的学习系统是较有影响的工作 [Winston, 1975]。

第三个阶段, 开始于 20 世纪 70 年代中期, 研究活动日趋兴旺, 各种学习方法不断推出, 实验系统大量涌现, 1980 年在卡内基·梅隆大学(CMU) 召开第一届机器学习专题研讨会, 标志着机器学习正式成为人工智能的一个独立研究领域。伴随专家系统和知识工程的热潮, 人们普遍认识到, 为了学到新知识, 学习系统必须拥有大量的知识。较为成功的学习系统和算法有 AQ 系列 ([Michalski, 1980; Michalski et al, 1983]), ID3 ([Quinlan,

1986a, 1986b]), TEX([Mitchell, 1982; Mitchell et al, 1983]), ACT 系统([Anderson, 1983, 1986]), AM 和 EURISKO([Lenat, 1982, 1983a, 1983b, 1983c]), BACON 系统([Langley et al, 1983, 1986]), SOAR([Rosenbloom et al, 1986; Laird et al, 1986, 1987])等。还创办了国际学术刊物《机器学习》(Machine Learning), 每年举行国际机器学习会议及各种专题性、地域性会议。

从 20 世纪 80 年代中后期到现在, 可以认为机器学习研究进入一个新阶段, 已经趋向成熟。首先是神经网络的复苏, 带动着各种非符号学习方法与符号学习并驾齐驱, 并且已超越 AI 研究范围, 进入到自动化及模式识别等领域, 掀起一场联结主义的热潮; 其次在已有学习方法的研究基础上, 开始考虑对多种方法加以集成, 多策略学习得到迅速发展; 多种学习方法和系统已经走出学术杂志和实验室, 进入实际应用领域。学习的计算理论开始对机器的学习能力进行形式化的分析和总结, 得到一系列理论成果, 如 PAC Learning 等。在各种重要的 AI 会议中, 关于机器学习的文章已占有较多数量和篇幅, 成为 AI 的主流研究领域之一。

### 1.1.3 机器学习方法

学习系统有各种分类标准, 如按基本策略, 或知识表示, 或应用领域, 或学习方法或结果层次等。这里介绍从学习推理理论来对学习过程分类, 见图 1.3[卢美律等, 1995]。

#### 1.1.3.1 归纳学习

综合性的归纳学习(Inductive learning)亦可称作经验学习(Empirical learning)、基于相似性的学习(Similarity-based learning 简记 SBL)、例子学习(Learning from examples)等。

在例子空间中, 给出关于某个概念的正例和反例集合, 归纳学习系统使用一般化(generalization)和特殊化(Specialization)操作在空间中进行搜索, 并根据学习倾向(Bias)进行抉择, 获得该概念的一般化描述, 这是一个从特殊实例到一般规律的推理过程, 其哲学和数学基础是源自 Bacon 归纳法以来的归纳理论。

归纳学习是发展最成熟的学习方法之一, 拥有形式化理论模型, 应用已达到商品化。代表性的方法和系统有版本空间法(Version Space)([Mitchell, 1982]), 决策树(ID3)([Quinlan, 1986a, 1986b]), AQ 算法([Michalski, 1980]), BACON 系统([Langley et al, 1983]), 概念聚类([Michalski et al, 1983; Stepp et al, 1986; Gennari et al, 1989])等。

#### 1.1.3.2 分析式学习

分析式学习(A analytical learning)的特点是运用大量领域知识来驱动学习进程。实例不再单纯是无关特征集合, 而是在已有背景知识环境下得到考察, 学习结果也消除了倾向的影响。主要方法有类比学习(Learning by Analogy)、示例学习(Case-based learning)和解释学习等。

类比是在具有相似性质的领域间发现知识的对应关系, 对于未知或不熟悉的领域中的概念, 运用类比, 借助于相似的、熟悉的领域中的概念进行学习, 称为类比学习。由于环境提供的信息仅和一个类似的问题有关, 学习单元必须发现其中的相似性, 并形成类似的规则和假设以解决当前问题。类比学习需要归纳和演绎推理: 寻找共同的子结构需要归纳

推理，而进行类比映射则是一演绎过程。典型类比系统模型有转换类比(Transformational Analogy) ([Carbonell, 1983])/派生类比(Derivational Analogy) ([Carbonell, 1986])/NLAG 系统([Greiner, 1989])等。类比包含形象思维和创造性成分,增大了实用系统的研究难度,同时也蕴含着推动机器学习和人工智能研究进展的潜力。

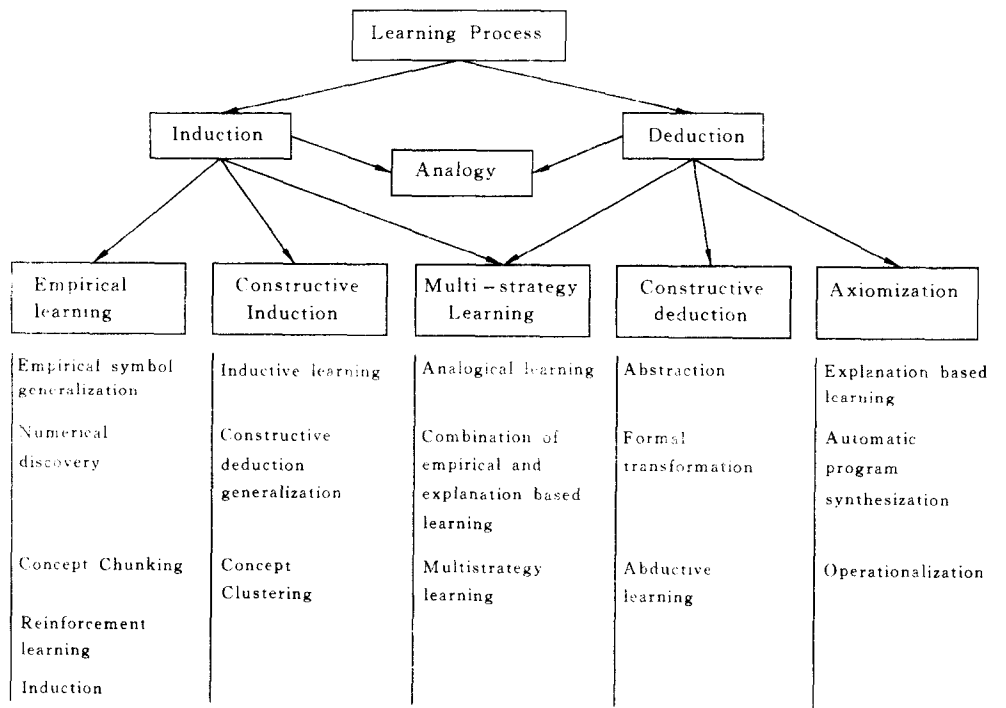


图 1.3 学习推理理论对于学习过程的分类

示例学习选择先前有用的经验片断作为示例,加以记忆和索引,建立示例库;在解决新问题时,在示例库中查询与当前任务比较类似的示例来加以匹配,借助它们所提供的信息进行任务求解;任务求解成功后,若认为该经验(任务目标及解法)有助于未来其它任务的解决,则将其作为新的示例加以保存[Bareiss et al, 1987; Porter et al, 1990]。示例学习系统的运行流程如图 1.4 所示。

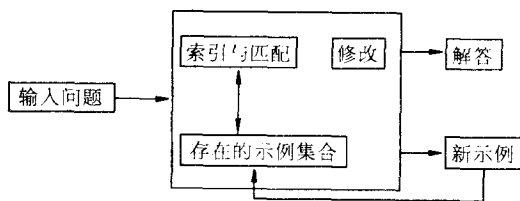


图 1.4 示例学习系统的运行流程[Barletta, 1993]

Schank 首先提出示例学习的思想[Schank, 1982]。1983 年 Kolonder 的 Cyrus 是第一个可在计算机上运行的示例学习系统;随后在不同领域中又研制出许多系统[李宝东,

1992; 卢美律等,1995]。示例学习方法的技术关键是,如何有效地选择和存储新的有用示例,为示例库建立便于查询的索引机制和恰当的匹配标准,以使相似示例的查找快速、方便和准确。

### 1.1.3.3 人工神经网络和遗传算法

神经网络和遗传算法(Genetic Algorithm)属于非符号学习范畴,不是在形式符号层次,而主要是在数值层次体现学习行为。

人工神经网络又称为联结学习(Connectionist learning)机制,是由大量简单处理单元组成的非线性动力系统,它是借助动物神经生理科学的研究成果,假定人脑神经网络加工、记忆信息的机理,利用计算机系统加以模拟。具体讲,人工神经网络是一组相互关联的神经元结点,每个结点都有其输入端、内部记忆状态和输出端,输出和内部状态由输入决定,而输入则由与之相连的若干结点的输出及其联结权值决定。神经网络的行为就取决于这些权值和输入-输出函数。给定网络的初始状态,系统经过计算最终达到平衡状态,完成一次问题求解。学习过程即调整结点间联结的权值,以期得到规定的输出。

神经网络的主要特点有:①大规模并行处理;②强健的分布式存储;③自适应、自组织的学习过程[杨行竣等,1992]。

主要研究成果有 MP 模型、感知机(Perceptron)、Hebb 假设、自适应线性网络 Adaline、连接网络模型、Boltzmann 机和多层网络学习算法、PDP 理论和多层网络误差反向传播算法(BP 算法)等。

遗传算法或称基因算法,源于对生物繁殖中的诱变原理和达尔文自然选择原理的计算机模拟,是由 Holland 于 1975 年首次提出[Holland,1986; Goldberg,1989]。它根据解决问题的规则在应用中效果的好坏和大小,进行优胜劣汰的繁殖:把参加每一轮竞争的规则群体(Population)看作一代,在同一代中选择表现优异者进行杂交,形成新一代规则;如此一代代地更新,以期获得一组高质量的规则。规则为适应应用环境而作的一代代更新,犹如生物为适应自然而更新后代。其运行流程为:[吴轶华,1991a]

- (1) 建立初始群体;
- (2) 评价群体中的个体;
- (3) 利用一组遗传算子:复制(Reproduction)、杂交(Crossover)、变异(Mutation)或倒向(Inversion),创建新一代群体,保持个体总数不变;
- (4) 评价新的个体;
- (5) 如果超过时间界限,或个体产生的迭代次数超过预设常数  $n$ ,则输出评价值最高的个体,否则转(3)。

分类器系统(Classifier System)是发展较为成熟的基于遗传的学习方法之一。一般结构如图 1.5 所示。有三个主要部分:

- (1) 规则发现 使用基因算法进行变形,生成新的行为规则;
- (2) 信度调整 利用桶队算法(Bucket Algorithm)调整规则的可信度,为规则评定级别和加(减)权值;
- (3) 执行部分 即规则和信息系统,具体负责问题求解。由外界给出一组消息,其形式为  $01\#$  ( $\#$ 代表 0 或 1 构成的串),系统检查输入消息是否与输入前提匹配,若匹配则

输出规则的结论消息。

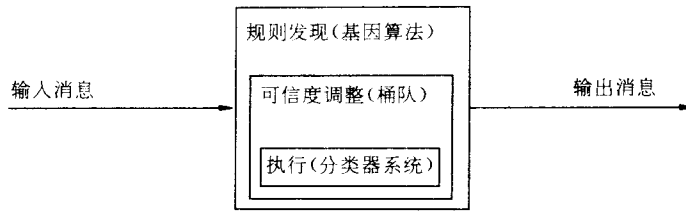


图 1.5 分类器系统的一般结构[吴轶华,1991b]

## 1.1.4 学习方法和系统的评价

### 1.1.4.1 衡量学习系统的标准

虽然还没有学习的形式化定义及定量评价手段,但人们在 ML 研究中,根据经验总结出 ML 系统应当遵循的若干标准,定性地给出 ML 系统的必要条件。大致说来,有以下几个方面。

#### (1) 有效性与效用性(Effectiveness and utility)

在学习的各种定义中,都强调在学习过程发生后,系统的性能应有所改善。对于 AI 系统,具体体现在使系统的问题求解能力(Ability)或效率(Efficiency)提高,这是学习的本质特征之一,是所有学习系统必须具备的。

要达到学习的有效性,就得考虑学习和问题求解如何结合,融为一体,这是 ML 研究中重要的课题。

有效性和效用性的含义基本相同,只是在若干场合,效用性的范围更广一些。

#### (2) 新知识的获得(Ability to get knowledge)

从无知到已知,从少知到多知,是学习的古典定义。在上述各定义中,也都强调系统内部变化的必然性。性能改善为果,系统内部变化、获得知识层次(Knowledge level)上的新知识则是因,是学习活动的内在表现。这在关于学习的定义(3)中反映得最明显。至于如何获得新知识,新知识表现为什么形式,则观点各异,并导致各种 ML 方法的出现,下节还要介绍。

上述内容是学习的两种本质属性。为了充分接近人的学习能力,为实现真正实用的 ML 系统,为评价具体系统的优劣,还应考虑下面几项重要的、现实性的度量。

#### (3) 正确性(Validity)

学习的结果必须是正确的。受认识活动的客观限制,一次学习活动的结果不可能达到完全正确。但总体地、长期地看,学习应该向正确结果收敛,至少是逼近。如果学习结果是可疑的,说明系统不可靠、不强健,无法达到改善性能的目的。对学习结果常用的验证手段,有实验、计算、逻辑推演等。

#### (4) 抽象性(Abstraction level)

学习决不仅仅是就事论事和对所经历的简单记忆与再现。学习是构造性、创造性的活动,应举一反三、见叶知秋,由数据和经验上升到理论,涉及不同抽象层次,学习越来越抽

象的事物,是人类学习的固有能力,也是研究学习机制的关键性问题。

#### (5) 通用性(Generality)

同 AI 相似,ML 的最终目的也有两个:研究人类的学习机制和建造高效、通用的 ML 系统。基于环境、基于任务等 ML 研究固然重要,可望尽早投入实际应用,为进一步发展提供经验,但通用性也是 ML 系统能力不可或缺的重要标准。在保证效率的前提下,通用性越强,学习的范围越广。

在知识表示、学习算法、应用领域等方面,应该尽量提高通用性。另一方面,取长补短,研究集成多种 ML 方法的混合系统,也是通用性提出的要求。

#### (6) 可理解性和表示(Understandability and representation)

学习的过程和结果必须以人和机器都能理解的方式表示,这是人们对 AI 系统所施加的传统要求。研究 ML 的目的就是理解 ML,运用 ML。如果学习过程无法为人们所理解,就无法实现人机间的交流,也不能验证学习结果的正确性,更无法加以有效的利用。定义(3)明确地以可表示性作为 ML 定义的基本内容,足见其重要性。

可理解性问题应从两方面来看。一方面,学以致用,ML 系统的算法和结果应该易理解、易表示、易应用。另一方面,人们尚未完全理解自身的学习机制,这说明人的理解能力也还有待提高。人工神经网络(Artificial Neural Network,简称 ANN)对 AI 的挑战就是一例。ANN 的学习机制有其独特的内在规律,在传统符号意义上是不可表达的。因此,要探究 ANN 与符号方法的关系,使 ANN 易于理解和运用;同时,也要提高我们的认识能力,弄清 ANN 的“表示”的含义。

#### (7) 强健性(Robustness)

ML 要从实验室中的游戏问题(Toy problem)转向实际应用领域,就要和模糊、不完全、不正确的外部环境打交道。这要求 ML 系统不仅能在正常状态下运行,也能在异常状态下正常地工作,至少不应崩溃或产生破坏性结果。所以,为了消除干扰、噪音和不可预料的故障,我们要求 ML 方法和系统具有强健性。

#### (8) 增量性(Increment)

增量式学习,是指 ML 系统在处理新的训练例时,能够只对原学习结果中与新例子有关的部分进行方便的增加、修改或删除操作,与之无关的部分则不被触及。随着学习过程的不断进行,学习结果越来越多,若其中与新输入的训练例无关的部分能免于处理,可以显著降低学习的花费,这正是增量式学习的魅力所在。

非增量式学习,每当输入新训练例时,需对新、旧训练例集合的并集进行处理,才能获得结果。伴随训练例不断输入,每次学习需处理的例子个数单调上升,花费也单调增加,最后无法忍受。所以,实际的 ML 系统,都应是增量式的。

#### (9) 自适应性(Adaptation)

自适应学习,是指 ML 系统在进行第  $n+1$  次学习时,以前  $n$  次学习的结果  $R_n$  和本次训练集合  $I_{n+1}$  作为输入,学到新结果  $R_{n+1}$ 。

图 1.6 为自适应学习的示意图。

显然,自适应学习必然是增量式的。只是

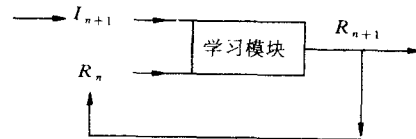


图 1.6 自适应学习过程



一般认为,增量式学习注重对学习结果中无关部分的不处理,目的是减少花费。而自适应体现了反馈的观点,学习结果反作用于学习过程,对低质量训练例的干扰的抵抗性增强,加速学习过程向平衡状态发展,是较为高级的学习。

#### (10) 自治性(Autonomy)

自治性是与依赖性相对的概念。只依赖经验和预备知识的自动、高效、通用的 ML 系统,至少目前是不可能的,必须有教师(系统建造者、领域专家或用户)的示教、解疑(离线的程序设计或在线的人机交互)。人类也同样不能够无师自通。但人在学习过程中,同教师的交互内容会逐渐发生质的变化。开始阶段所知甚少,频频提问,问题浅显、幼稚、缺乏目的性;随着知识的增加,问题变得抽象、关键而有针对性,交互次数减少,自治性增强,教师的负担越来越小,体现了学习的“进化性”。ML 系统也不例外,根据交互的次数和提问性质(即自治性)也能衡量 ML 系统达到的程度和能力。对 ML 系统而言,不实现自治性的增长,不会达到与人相似的学习能力。自治性是比自适应性更高的要求。

#### (11) 对知识的利用(Dependency on knowledge)

AI 和 ML 的研究进程,都经历了从寻求通用机制到对领域知识、常识知识和元知识的利用这一转变,重新肯定“知识就是力量”。只依靠一两种通用机制,没有足够数量的预备知识做前提,是不可能实现可与人类相比拟的 AI 和 ML 系统的,对已有的知识利用得越好,系统越有前途。但如何利用知识,基于知识的系统如何实现“自知”([GEB:一条永恒的金带]),知方能学与学方能知这对矛盾如何处理,都是有待解决的问题。

#### (12) 集成性(Integration)

集成性,就是度量 ML 方法是否开放,是否易于与其它 ML 方法及 AI 中的各种方法相集成。现有的 ML 方法,各具特色,互有长短,都不能包打天下。这说明各种 ML 方法揭示了整体认知学习能力的不同侧面。在无法肯定统一的学习机制存在之前,将各种 ML 方法互相融合,取长补短,获得整体性能的改善,亦有重要的理论和实用价值。作为智能体本质特征的学习能力,和 AI 有相关的困难和希望,集成性即为 AI 中综合集成观点在 ML 研究中的体现。ML 方法集成性越高,对 ML 和 AI 整体研究的贡献越大。

#### (13) 形式化水平(Level of Formality)

ML 方法中基本概念和机制的形式化水平,反映着该方法研究的成熟程度。对 ML 方法进行形式化,能够揭示出方法的本质思想,给出其能力水平和特点,是自身成立和广泛应用的前提。上面给出了衡量 ML 方法的若干建议和标准。对具体的 ML 方法而言,有些标准也许并不恰当,甚至在一定程度上相互矛盾。但在整体上综合评判,希望能促进 ML 方法间的客观比较,提出改进方向,为达到揭示学习本质的最终目的做出贡献。

### 1.1.4.2 ML 方法的比较

本节运用上节中给出的评价标准,对各种 ML 方法进行比较和初步分析,以便对 ML 领域研究现状形成概观。当然,ML 方法的思想、目的、应用领域和使用技术各异,互有优劣,不必拘泥于特定标准加以苛求。比较结果见表 1.1。