归纳学习

——算法理论应用

洪家荣 著

科学出版社

归 纳 学 习

算法 理论 应用

洪家荣 著

科 学 出 版 社
1997

内 容 简 介

机器学习，目前是人工智能的热门学科，归纳学习是机器学习最重要、最核心，也最成熟的一个分支。归纳学习旨在从大量的经验数据中归纳抽取出一般的规则和模式，因而成为知识获取的主要手段，在专家系统、模式识别、图像处理、语音识别等领域都有重要应用。本书概述了归纳学习的主要内容，包括主要学习算法、学习的计算理论，以及学习的各种应用。全书以作者的研究工作为基础，自成体系，内容新颖、丰富，算法描述清晰，理论推导严谨，应用范围广泛，是机器学习方面少有的专著之一。

本书可作为计算机科学专业研究生、本科生有关课程的教材或教学参考书，也可供相关专业研究人员参考。

归 纳 学 习

算法·理论·应用

洪家荣 著

责任编辑 马长芳

科学出版社出版
北京东黄城根北街 16 号
邮政编码：100717

中国科学院印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

*

1997 年 9 月第 一 版 开本：850×1168 1/32
1997 年 9 月第 次印刷 印张：7 1/4
印数：1— 1-00 字数：129 400

ISBN 7-03-005898-4/TP・779

定价：18.00元

前　言

机器学习(machine learning)当前已成为人工智能的一个热门学科。归纳学习(inductive learning)旨在从大量的经验数据中归纳抽取一般的判定规则和模式，它是机器学习最核心、最成熟的分支。归纳学习由于依赖于经验数据，因此又叫经验学习(empirical learning)；由于归纳依赖于数据间的相似性，所以它也叫做基于相似性的学习(similarity based learning)。归纳学习根据有无导师指导，又分为有导师学习(supervised learning)和无导师学习(unsupervised learning)。有导师学习是事先将训练例子(经验数据)分类，因此通常叫示例学习(learning from examples)；由于它产生规则，因而也叫概念学习(concept learning)。无导师学习事先不知道训练例子的分类，它包括概念聚类(conceptual clustering)和机器发现(machine discovery)。神经网络(neural networks)本质上也是一种示例学习。因此为区别起见，神经网络也叫联结学习(connectionist learning)。而上述各种学习方法统称符号学习(symbolic learning)。归纳学习的另一个领域是学习的计算理论(computational theory of learning)，它包括两个方向，一个是传统的算法复杂性分析，另一个是近年出现的概率近似正确性学习研究(probably approximately correct learning)或计算学习理论(computational learning theory)。归纳学习的早期研究主要集中在算法研究，近年来它的应用研究受到普遍的关注。上述各项构成了本书的主要内容。

本书的特点是以作者的研究工作为基础，内容新颖、丰富、自成体系，特别在学习算法设计与实现、学习的计算理论的研究及归纳学习的实际应用等诸方面都有独特的见解。本书内容比较新颖和完整。

机器学习的发展大致分为三个阶段：五六十年代的探索阶段；七十年代的发展阶段；八九十年代的鼎盛阶段。在第一阶段，机器学习受神经生理学、生理学和生物学的影响，研究主要集中在对神经元的计算模拟，其中最有代表性的工作是 Rosenblatt 的感知机(PERCEPTRON)，其次是 Friedberg 等模拟随机突变和自然选择过程的程序。此外，Hunt 等的决策树归纳程序 CLS 是早期的符号学习研究。然而，自从 1969 年 Minsky 和 Papert 揭示 PERCEPTRON 的严重局限性以后，神经元的人工模拟研究受到了打击而转入低潮。这个阶段较有影响的工作是 Samuel 的国际象棋程序，该程序通过学习能够达到大师级水平。

第二个阶段有影响的工作有 Winston 积木世界学习系统(1970 年)，该系统能够从例子学习积木世界的结构；Michalski 基于逻辑的归纳学习系统 AQVAL(1972)，以及 Michalski 和 Chilausky 的 AQ11(1980)，AQ11 产生的大豆病诊断规则的诊断正确率甚至超过大豆病理专家的水平，在当时曾引起很大的反响；Quinlan 的决策树归纳程序 ID3(1983)，它用来学习象棋规则获得很大的成功；此外还有 Mitchell 的 Version Space(1982)，它使用候选消去算法进行归纳学习。

第三个阶段各种学习算法均得到极大发展，理论研究与应用研究也有新的突破。在归纳学习方面的发展尤其突出显著。(1)在学习算法研究方面，Michalski 和作者等将 AQ11 扩充为一个多功能通用学习系统 AQ15(1986)，Quinlan 在其 ID3 系统中使用了熵，从而使决策树归纳得到很大的改进；近来作者提出的示例学习的扩张矩阵方法、系统 AE1(1985)，以及吴信东的改进算法 HCV(1992)等越来越引起普遍的关注。Michalski 和 Stepp 的概念聚类系统 CLUSTER/2(1983)和 Langley 等的科学发现系统 BACON.4(1987)开辟了无导师学习的两个重要领域；此外，知识发现(knowledge discovery)研究近年来得到飞速的发展。联结学习或神经网络研究在消沉了 20 年以后于 80 年代中期又蓬勃发展起来，其中 Rumelhart 等的 BP 网络有很大影响(1986)，因为它提供了一

个训练多层网络的实际可行的方法，从而克服了 PERCEPTRON 的大部分局限性。(2)在学习的计算理论研究方面，1984 年美国学者 Valiant 提出基于概率近似正确性的学习理论，简称 PAC 学习，也叫计算学习理论(COLT)，从而对布尔函数的一些特殊子类的可学习性进行了探讨。近几年 PAC 学习研究在国际上掀起了一阵热潮。然而，计算机界最常用的两类布尔表达式——析取范式 DNF 和合取范式 CNF 是否是 PAC 可学习的问题，却一直悬而未决。在本书中，作者将证明 DNF 和 CNF 都是 PAC 不可学习的，从而揭示出 PAC 学习的局限性。此外，作者在 PAC 学习的两篇奠基性论文中从理论角度指出了差错。因此，学习的传统计算复杂性理论引起了人们的重视，这方面的早期工作有作者证明了的学习最小 DNF 规则是 NP 难题(1985)。本书将给出所有知名学习算法的计算复杂性分析及其启发式算法。(3)近年来应用研究成为机器学习的焦点。例如，AQ15 已经达到实用化，并成功地应用于三种疾病的诊断；ID3 的应用越来越广泛；科学发现已能解决实际领域的问题，特别是作者成功地应用于彩色匹配；数据库中的知识发现已面向实际应用；而联结学习的实际应用获得更大的成功，在模式识别、图像处理及语音识别等领域被广泛应用。估计，机器学习的应用研究仍然是未来 10 年的研究热点。

本书共分四章，其中第 2，6，7，8 和 9 等五部分是为了系统和完整而引用的经典之作。其余十二部分基本上为作者的研究成果。第一章主要介绍各种有导师学习算法，包括 AQ11，AQ15，AE1，HCV 及作者提出的 GS 及快速覆盖算法 FCV，以及著名的 ID3。第二章介绍学习的计算理论，包括学习的形式理论、上述各种学习算法的计算复杂性分析及 Valiant 的 PAC 学习理论，给出了一般 DNF 和 CNF 表达式 PAC 不可学习的证明，指出了 PAC 学习缺乏实际意义的这一固有缺欠。第三章介绍无导师学习算法，包括 Michalski 和 Stepp 的概念聚类系统 CLUSTER/2，Fisher 的渐近概类系统 COBWEB(1987)，Langley 的科学发现系统 BACON，作者等提出的基于逆归函数的科学发现方法(1994)，以及数据库中

知识发现算法 KD3(1989)。第四章介绍归纳学习方面的应用研究,包括基于科学发现的彩色匹配系统、专家系统实时化工具 THOUGHT、OX 报表自动录入系统、三角割分的模拟退火方法、彩色抖动的模拟退火方法、实用化自动知识获取系统 PKAS,以及 AQ15 在三个医学领域的应用。

目　录

前　言
第一章　有导师学习算法 …… 1
1　覆盖算法 …… 1
1.1　覆盖算法的例子 …… 1
1.2　基本概念 …… 3
1.3　星算法与 AQ11 …… 4
1.4　AQ15 …… 7
1.5　扩张矩阵算法与 AE1 …… 11
1.6　广义扩张矩阵与 AE9 …… 15
1.7　扩张矩阵的启发式算法与 HCV …… 21
1.8　快速覆盖算法 FCV …… 24
1.9　一个简单的贪心算法 GS …… 30
2　分治算法 …… 33
2.1　CLS 系统 …… 33
2.2　ID3 算法 …… 34
2.3　ID3 的某些改进算法 …… 37
3　联接学习 …… 41
第二章　学习理论 …… 43
4　学习的形式理论 …… 43
5　学习的计算理论 …… 45
5.1　最优覆盖问题 …… 46
5.2　最一般复合问题 …… 52
5.3　最小决策树问题 …… 54
5.4　精确训练神经网络问题 …… 56
5.5　最小属性子集问题 …… 57
6　计算学习理论 …… 58

6.1 PAC 学习 …… 59
6.2 Occam 算法 …… 63
6.3 VC-Dimension …… 64
6.4 计算学习理论的局限性 …… 65
6.5 COLT 的一个公开问题的解决 …… 68
6.6 某些经典文献中的理论错误 …… 77
6.7 k-term DNF 是可学习的 …… 78
第三章 无导师学习算法 …… 81
7 概念聚类 …… 81
7.1 动态聚类 …… 81
7.2 概念聚类 …… 82
8 概念形成 …… 84
8.1 COBWEB 算法 …… 85
8.2 例子 …… 87
8.3 评价函数 …… 88
9 科学发现 …… 89
9.1 BACON 系统 …… 89
9.2 科学发现的递归函数法 …… 91
10 知识发现 …… 96
10.1 知识发现的计算模型 …… 96
10.2 应用举例 …… 99
第四章 归纳学习的应用 …… 103
11 科学发现用于彩色匹配 …… 103
11.1 色彩的基本关系 …… 104
11.2 基于科学发现的 RGB-CMYK 空间变换 …… 104
12 专家系统实时化工具 THOUGHT …… 107
13 OX 报表自动录入系统 …… 113
13.1 表格识别与理解子系统 …… 113
13.2 数字识别子系统 …… 116
13.3 印刷汉字识别子系统 …… 121
14 三角剖分的模拟退火方法 …… 124
14.1 组合优化问题的模拟退火解法 …… 125

14.2 模拟退火算法的实现 …… 127
14.3 模拟退火解三角剖分问题 …… 128
14.4 同现有三角剖分算法比较 …… 133
15 基于模拟退火的图像输出抖动模式的研究 …… 134
15.1 基本概念 …… 134
15.2 模拟退火求解抖动模式 …… 136
15.3 评价函数及其改进 …… 141
16 一个实用化知识获取系统 PKAS …… 145
16.1 贝叶斯分类 …… 145
16.2 GS 算法 …… 147
16.3 检测 …… 147
16.4 实例 …… 148
16.5 实验结果及结论 …… 150
17 AQ15 应用于医疗诊断问题 …… 151
参考文献 …… 152

第一章　有导师学习算法

在这一章里，我们将介绍各种示例学习算法。这些学习算法主要分为两大类：覆盖算法(covering algorithms)和分治算法(divide-and-conquer algorithms)。覆盖算法归纳生成规则，一般是析取范式(DNF)；分治算法归纳生成树状结构，叫做决策树(decision trees)。知名的覆盖算法有：AQ11(Michalski Chilausky，1980)及其扩充 AQ11(Michalski，Hong 等，1986)，AE1(Hong，1985)及其改进 AE9(赵美德、洪家荣，1995)、HCV(Wu，1992)，FCV1(陈彬、洪家荣，1996)，GS(洪家荣，1987)。知名的分治算法有：CLS(Hunt 等，1966)及其改进 ID3(Quinlan，1983)、C4.5(Quinlan，1994)。近年来，国外一些学者主张将以作者的扩张矩阵算法 AE1 为代表的一类算法归为新启发式算法(Wu，1994)。

1　覆盖算法

1.1　覆盖算法的例子

假定有两组人，其中每个人具有如下三种特征：(1) 身材：高或矮；(2)发色：金色、黑色或红色；(3) 眼睛(颜色)：蓝色、黑色或灰色。即每个人以一个向量(身材，发色，眼睛)来表征。表 1.1.1 给出了这两组人特征向量(叫做例子)的集合。

示例式学习的目的就是从两类(或多类)例子的集合中找出描述其中一类而排除另一类(或其余类)的一般规则。例如，对上述例子运行系统 AE1，我们可以得到如下两条规则：

第 1 组：[发色＝金色∨红色][眼睛＝蓝色∨灰色]

第 2 组：[发色＝黑色]∨[眼睛＝黑色]

第 1 组的描述规则是说，凡具有金色或红色头发并且同时具有蓝

色或灰色眼睛的人属于第1组；第2组的描述规则是说，凡具有黑色头发或黑色眼睛的人属于第2组。

表 1.1.1　两组人(例子)的集合

组　次	变元 / 例	身　材	发　色	眼　睛
第1组	1	矮	金　色	蓝　色
	2	高	红　色	蓝　色
	3	高	金　色	蓝　色
	4	矮	金　色	灰　色
第2组	1	高	金　色	黑　色
	2	矮	黑　色	蓝　色
	3	高	黑　色	蓝　色
	4	高	黑　色	灰　色
	5	矮	金　色	黑　色

以下为讨论方便，我们把一个离散符号集映射到非负整数集上，并且把正、反例集 PE 与 NE 列成矩阵的形式并仍记为 PE 与 NE，如表 1.1.2 所示。

表 1.1.2　两组人的数值表示

(a)人的特征的数值表示

v_i \ x_j	身材	发色	眼睛
0	矮	金色	蓝色
1	高	黑色	黑色
2		红色	灰色

(b)正例矩阵与反例矩阵

k	x_1	x_2	x_3	k	x_1	x_2	x_3
1	0	0	0	1	1	0	1
2	1	2	0	2	0	1	0
3	1	0	0	3	1	1	0
4	0	0	2	4	1	1	2
				5	0	0	1
	PE				NE		

这个例子表明，示例式学习系统（如 AE1）能够从大量例子（如医学上的病例）归纳出较少的描述规则（如诊断），从而可以实现知识的自动获取。

1.2 基本概念

设 E 是一个 n 维离散符号的有穷向量空间，即 $E=D_1\times D_2\times\cdots\times D_n$，其中 D_j 是有穷离散符号集，$j\in N$，$N=\{1,2,\cdots,n\}$ 为变元下标集。PE 和 NE 是 E 的子集并分别叫做正例集与反例集。D_j 的子集 $A=\{v_1,v_2,\cdots,v_t\}$ 可以表示成内部析取 $v_1\vee v_2\vee\cdots\vee v_t$，$E$ 中的元素 e 叫做一个例子，记为 $e=\langle v_1,\cdots,v_n\rangle$，其中 $v_j\in D_j$，$j\in N$。

定义 1.2.1 选择子（selector）是形为 $[x_j\ \#\ A_j]$ 的关系语句，其中 x_j 是第 j 个变元，$A_j\subseteq D_j$，关系 $\#\in\{=,\neq,>,\geqslant,<,\leqslant\}$。公式或复合（complex）为选择子的合取式，记为 $\bigwedge\limits_{j\in J}[x_j=A_j]$，或补形复合 $\bigwedge\limits_{j\in J}[x_j\neq A_j]$，其中 $\wedge$ 为合取并往往省略，$J\subseteq N$。注意，

$$[x_j\neq A_j]\equiv[x_j=D_j\backslash A_j]$$

定义 1.2.2 已知例子 $e=\langle v_1,\cdots,v_n\rangle$，选择子 $S=[x_j\neq A_j]$ 及公式

$$L=\bigwedge_{j\in J}[x_j\neq A_j]$$

e 满足选择子 S 当且仅当 $v_j\notin A_j$；e 满足公式 L 当且仅当 e 满足 L 的每一个选择子，即对所有 $j\in J$，$v_j\notin A_j$，e 满足 $S(L)$ 也叫做 $S(L)$ 覆盖 e。

给定一个正例集 $PE=\{e_1^+,\cdots,e_k^+\}$ 及反例集 $NE=\{e_1^-,\cdots,e_m^-\}$，其中 $e_i^*=\langle v_{i1}^*,\cdots,v_{in}^*\rangle$，而 $*\in\{+,-\}$。

定义 1.2.3 正例 e^+ 在反例 e_i^- 背景下满足公式 L 当且仅当 e^+ 满足 L，但 e_i^- 不满足 L。e^+ 在反例集 NE 背景下满足 L 当且仅当 e^+ 在每个反例 $e_i^-\in NE$ 背景下满足 L，$1\leqslant i\leqslant m$。公式的集合或析取式 COV 叫做正例集 PE 在反例集 NE 背景下的一个规则或覆盖当且仅当 PE 中的任何一个正例都在 NE 背景下至少满足

COV 中的一个公式。

定义 1.2.4 已知正例集 PE、反例集 NE，学习算法 A。算法 A 归纳产生覆盖 PE 且排除 NE 的规则记为 Cover(PE, NE, A)，有时简记为 Cover(PE, NE)。

1.3 星算法与 AQ11

最早的覆盖算法是 Michalski 的 A^q，其意为准优算法(quasi-optimal algorithm)，A^q 算法的核心是星(star)算法，其定义如下。

定义 1.3.1 已知正例集 $PE=\{e_1^+,\cdots,e_n^+\}$ 及反例集 $NE=\{e_1^-,\cdots,e_m^-\}$，其中 $e_i^*=\langle v_{i1}^*,\cdots,v_{in}^*\rangle$，而 $*\in\{+,-\}$。一个正例 e_i^+ 在反例 e_j^- 背景下的星记为 $G(e_i^+|e_j^-)$ 是所有覆盖 e_i^+ 而排除 e_j^- 的极大复合(maximal complex)的集合。这里一个极大复合是除覆盖 e_i^+ 排除 e_j^- 之外覆盖最多数目的其他正例的复合。在星 $G(e_i^+|e_j^-)$ 中，正例 e_i^+ 叫做种子。正例 e_i^+ 在反例集 NE 背景下的星记为 $G(e_i^+|NE)$ 是一切覆盖种子 e_i^+ 且排除 NE 中的所有反例的极大复合的集合。

下列引理表明如何构造星。

引理 1.3.2 $e_i^+=\langle v_{i1}^+,\cdots,v_{in}^+\rangle$ 在反例 $e_j^-=\langle v_{j1}^-,\cdots,v_{jn}^-\rangle$ 背景下的星 $G(e_i^+|e_j^-)$ 是一逻辑公式 $e_i^+\wedge\neg e_j^-$ 即 $([x_1=v_{i1}^+]\wedge\cdots\wedge[x_n=v_{in}^+])\wedge\neg([x_1=v_{j1}^-]\wedge\cdots\wedge[x_n=v_{jn}^-])=([x_1=v_{i1}^+]\wedge\cdots\wedge[x_n=v_{in}^+])\wedge([x_1\neq v_{j1}^-]\vee\cdots\vee[x_n\neq v_{jn}^-])$ 展开为析取范式 DNF 中所有析取项中覆盖正例数最多的那些项的集合。e_i^+ 在 NE 背景下的星 $G(e_i^+|NE)$ 是逻辑公式 $e_i^+\wedge\neg NE=e_i^+\wedge\neg(e_1^-\vee\cdots\vee e_m^-)=e_i^+\wedge\neg e_1^-\wedge\cdots\wedge\neg e_m^-$ 展开式中那些极大复合的集合。

证明：由星的定义即得。

易见，星 $G(e_i^+|NE)$ 的展开式共有 n^m 个复合，其中 n 是变元数，m 是反例数。因此，直接构造星是不实际的。Michalski 在 A^q 中使用缩小的星(reduced star)。另一方面，在第二章我们将证明寻找一个极大复合的问题是 NP 难的。因此，只好使用启发式算法。

定义 1.3.3 e_i^+ 在反例 e_j^- 背景下的缩小的星记为 $G(e_i^+|e_j^-, MS)$，是在星 $G(e_i^+|e_j^-)$ 中选出 MS(Maxstar)个复合组成的集合。e_i^+ 在 NE 背景下的缩小的星记为 $G(e_i^+|NE, MS)$，是在星 $G(e_i^+|NE)$ 每次展开取 MS 个复合同下一次展开式进行逻辑乘而得到的复合的集合，其中 MS 是一个较小的整数。

我们以表 1.1.2 的正反例集为例，构造一个星 $G(e_1^+|NE)$。

$$G(e_1^+|NE) = e_1^+ \wedge \neg e_1^- \wedge \neg e_2^- \wedge \neg e_3^- \wedge \neg e_4^- \wedge \neg e_5^-$$

为此计算

$$[x_1=0][x_2=0][x_3=0] \wedge ([x_1 \neq 1] \vee [x_2 \neq 0] \vee [x_3 \neq 1]) \wedge ([x_1 \neq 0] \vee [x_2 \neq 1] \vee [x_3 \neq 0]) \wedge ([x_1 \neq 1] \vee [x_2 \neq 1] \vee [x_3 \neq 0]) \wedge ([x_1 \neq 1] \vee [x_2 \neq 1] \vee [x_3 \neq 2]) \wedge ([x_1 \neq 0] \vee [x_2 \neq 0] \vee [x_3 \neq 1])$$

$$= (e_1^+[x_1 \neq 1] \vee e_1^+[x_3 \neq 1]) \wedge ([x_1 \neq 0] \vee [x_2 \neq 1] \vee [x_3 \neq 0]) \wedge \neg e_3^- \wedge \neg e_4^- \wedge \neg e_5^-$$

$$= (e_1^+[x_1 \neq 1][x_2 \neq 1] \vee e_1^+[x_1 \neq 1][x_2 \neq 1]) \wedge ([x_1 \neq 1] \vee [x_2 \neq 1] \vee [x_3 \neq 0] \wedge \neg e_4^- \wedge \neg e_5^-$$

$$= (e_1^+[x_1 \neq 1][x_2 \neq 1] \vee \underline{e_1^+[x_3 \neq 1][x_2 \neq 1][x_1 \neq 1]} \vee e_1^-[x_3 \neq 1][x_2 \neq 1]) \wedge (\text{第二项为冗余项})([x_1 \neq 1] \vee [x_2 \neq 1] \vee [x_3 \neq 2]) \wedge \neg e_5^-$$

$$= (e_1^+[x_1 \neq 1][x_2 \neq 1] \vee \underline{e_1^+[x_1 \neq 1][x_2 \neq 1][x_3 \neq 2]} \vee \underline{e_1^+[x_3 \neq 1][x_2 \neq 1][x_1 \neq 1]} \vee e_1^+[x_3 \neq 1][x_2 \neq 1] \vee \underline{e_1^+[x_3 \neq 1][x_2 \neq 1][x_3 \neq 2]})(\text{划线者为冗余项而删除}) \wedge ([x_1 \neq 0] \vee [x_2 \neq 0] \vee [x_3 \neq 1])$$

$$= e_1^+[x_1 \neq 1][x_2 \neq 1][x_3 \neq 1] \vee e_1^+[x_3 \neq 1][x_2 \neq 1][x_1 \neq 0] \lessdot [x_1 \neq 1][x_2 \neq 1][x_3 \neq 1] \vee [x_3 \neq 1][x_2 \neq 1][x_1 \neq 0] = [x_2 \neq 1][x_3 \neq 1]$$

最后一次运算是删除合取项 e_1^+ 为了使产生的复合极大化，因将原来的公式概括化(generalization)，而不再保持逻辑等价关系，符号 $\leftarrow$ 表示"概括为"的意思。A^q 算法的归纳性正在于这种放松约束条件的运算，不然的话，产生的复合不能覆盖其他正例。

由于结果的两项覆盖的正例分别为 $\{e_1^+, e_4^+\}$ 与 $\{e_1^-, e_2^+, e_3^+, e_4^-\}$，因此 A^q 算法取第二项做为结果，即

$$[x_2 \neq 1][x_3 \neq 1] = [x_2 = 0,2][x_3 = 0,2]$$

转换为符号表示为[发色＝金色 V 红色][眼睛＝蓝色 V 灰色]，此同 1.1 节 $AE1$ 的结果。

为了直观地表示星，Michalski 使用图表示，如图 1.3.1 所示。

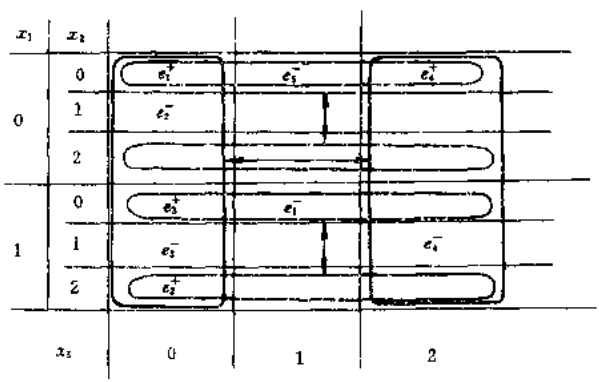

图 1.3.1 星的图表示，箭头连接的区域为复合[x_2=0,2]与[x_3=0,2]所覆盖的区域

从图 1.3.1 可见，种子 e_1^+ 既被横向矩形所覆盖，又被纵向矩形所覆盖，因此以该种子为中心构成一放射状的星，此即 Michalski 星的名字的来源。

基于构造星的 A^q 算法描述如下。

[A^q 算法]

已给正例集 $PE=\{e_1^+, \cdots, e_k^-\}$ 及反例集 $NE=\{e_1^-, \cdots, e_m^-\}$，$e_i^+$

$=\langle v_1^*,\cdots,v_m^*\rangle$，其中 $*\in\{+,-\}$。常数为 MS。

(1) Rule←ϕ。

(2) 如果 $PE=\phi$，则终止，返回规则 Rule；否则取一个种子 $e_l\in PE$，从 PE 中删去 e_l^+，即 $PE=PE-\{e_l^+\}$。

(3) 构造缩小的星 $G(e_l^+ \quad NE,MS)$，并根据 LEF 准则，将产生的复合按覆盖正例数的大小排序，并选出前 MS 个复合。

(4) 在最后产生的 MS 个复合中，选择覆盖正例最多的一个复合记为 cpx 放入 Rule 中，Rule←Rule$\cup\{$cpx$\}$。

(5) 从 PE 中删去被 cpx 覆盖的正例。

(6) 返回(2)。

[算法复杂性]

不难估计，生成一个复合需要计算量为

$$n^2+MS\times n\times(m-1)\simeq n\times(n+MS\times m)$$

其中，n 为属性数，m 为反例数，MS 为常数。生成一个规则需要计算量为

$$n\times(n+MS\times m)\times k=\Omega(n\times k\times m)=\Omega(n\times(k+m)^2)$$

其中 k 为正例数。因此，A^q 算法的计算复杂度大约为例子数的平方。

1.4 AQ15

由 Michalski 领导设计、但由作者实现的多功能学习系统 AQ15 是 AQ11 的实质改进和全面扩充，并成为目前覆盖算法的代表。

AQ15 的核心算法基本上延续 AQ11 的 A^q 算法，只是做了一些优化。AQ15 增加的主要功能有构造性学习、渐近学习和近似推理。

(1) 构造性学习模块。用户可以提供一些可供学习的背景知识。这些知识一般以算术表达式或逻辑规则的形式出现。AQ15 应用这些背景知识于例子的集合 PE 和 NE，产生原来没有的新概念(即属性或变元)，然后从这些新属性中筛选出一部分具有最强