示例式学习的两个方法及其应用

洪家荣 刘 宁 （哈尔滨工业大学计算机系）

摘 要 本文介绍示例式学习的两个主要方法：扩张矩阵法和决策树法，以及它们的主要应用。

随着专家系统研究的深入，知识获取问题已经成为构造专家系统的瓶颈。机器学习被认为是克服知识获取瓶颈的关键，它已成为人工智能研究的中心问题。机器学习已研究了若干学习方法[1]，如示例式学习(learning from example)、观察式学习(learning by observation)、发现式学习(learning by discovery)和类比式学习(learning by analogy)等。在这些学习方法中，示例式学习是最基本的学习方法，由于它能够从大量的例子中自动地归纳总结出规则，从具体的事例中抽象出概念，因此人们把它看作解决知识获取问题的重要途径，一直受到广泛而深入的研究。

示例式学习是从已知的正例集和反例集中，归纳出能描述正例集而排除反例集的一般描述规则。目前世界上已有一些示例学习方法和系统，其中最著名的是Michalski的AQ11[2]及其作者对它的扩充AQ15[3]、Quinlan的ID3[4]以及作者的GS[6]、AE1[7]及其扩充AE5[8]。本文介绍的扩张矩阵法是AQ15，GS，AE1，AE5的基础，决策树法是ID3的基础。作者在文[7]中于85年率先证明了最优示例式学习问题是NP难题，值得注意的是，三年以后，Haussler重复了作者相同的证明[5]，并被西方误认为首创者。直到最近一期《Artificial Intelligence》才声明确认作者的领先结果。

1. 扩张矩阵法及应用

本文引用文[8]的描述方式。

1.1 一个自动知识获取的例子

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。图1给出了这两组人特征向量（即例子）的集合。

组次	例 \ 变元	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

图1 两组人（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，对上述例子使用扩张矩阵法，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。

这个例子表明，示例式学习系统能够从大量例子（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 基本概念与理论

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$A=\{v_1, v_2, \cdots, v_l\}$可以表示成内部析取$v_1\vee v_2\vee\cdots\vee v_l$。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 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，$A_j\in D_j$，关系$\#\in\{=, \neq, >, \geqslant, <, \leqslant\}$。公式或复合(complex)为选择子的合取式，记为$\bigwedge_{j\in J}[x_j=A_j]$或补形复合$\bigwedge_{j\in J}[x_j\neq A_j]$，其中$\bigwedge$为合取（往往省略），$J\in N$。注意，$[x_j\neq A_j]\equiv[x_j=D_j\setminus A_j]$。

定义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。

给定一正例集$PE=\{e_1^+, \cdots, e_k^+\}$及反例集$NE=\{e_1^-, \cdots, e_m^-\}$，其中$e^*=\langle v_{i1}^*, \cdots, v_{in}^*\rangle$，且$*\in\{+, -\}$。

定义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。

1.3 扩张矩阵

不难看出，应用示例式学习获取知识就是寻找正例集在反例集背景下所满足的规则。下面的引理给出一个正例背景下满足一个公式的条件。

引理 已知正例$e^+=\langle v_1^+, \cdots, v_n^+\rangle$，$NE=\{e_1^-, \cdots, e_m^-\}$，其中$e_i^-=\langle v_{i1}^-, \cdots v_{in}^-\rangle$，及公式$L=\bigwedge_{j\in J}[x_j\neq A_j]$，则$e^+$在NE背景下满足公式L当且仅当

（i）对所有$j\in J$，$v_j^+\notin A_j$，

（ii）对所有i，$1\leqslant i\leqslant m$，至少存在一个$l\in J$，使得$v_{il}^-\in A_l$。

证 由定义2与定义3直接推出。

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，如图2所示。

定义4 已知$e^+=\langle v_1^+, \cdots, v_n^+\rangle$及反例矩阵NE。对每一$j\in N$，用"死元素"*对$v_j^+$在NE中第j列的所有出现做代换，这样得出的矩阵叫做正例e^+在反例NE背景下的扩张矩阵，记为$EM(e^+)$。

图3给出了图2(b)中的四个正例在五个反例背景下的扩张矩阵。

v \ x	身材	发色	眼睛
0	矮	金色	蓝色
1	高	黑色	黑色
2	高	红色	灰色

(a) 人的特征的数值表示

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

(b) 正例矩阵与反例矩阵

图 2 两组人的数值表示

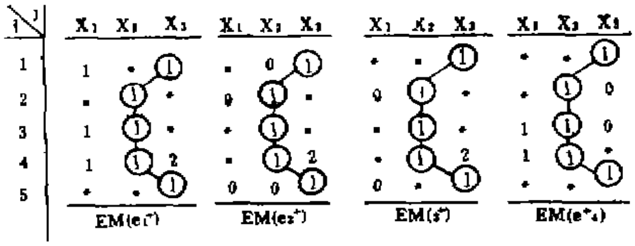


图3　正例在反例集背景下的扩张矩阵

定义5　在扩张矩阵EM(e^+)中，由分别来自不同行的m个非死元素组成的集合叫做一条路。在两个以上的扩张矩阵中，具有相同值的对应的非死元素叫做它们的公共路。具有公共路的扩张矩阵叫做相交的；否则叫做不相交的。

在图3中，每个扩张矩阵中画出的路就是它们的公共路。下面的定理给出了扩张矩阵的性质。

定理1　扩张矩阵EM(e^+)中的路同e^+在NE背景下所满足的公式一一对应。

证　（=>）

设$P=\langle r_{1j}, r_{2j}, \cdots, r_{mj}\rangle$是EM($e^+$)中的一条路，则公式

$$L=\bigwedge_{i=1}^{m}[x_j\neq r_{ij}]$$

在NE背景下覆盖e，事实上，由EN(e^+)的定义，e^+的所有元素均不在L中出现，因而根据定义2，L覆盖e^+；另一方面，这m个元素分别来自每一个反例，所以再根据定义2，L不覆盖NE中的任何一个反例。

（<=）设公式$L=\bigwedge_{j\in J}[x_j\neq A_j]$在背景NE下覆盖$e^+$。则对于每个i，$1\leqslant j\leqslant m$，EM($e^+$)中必有一个元素，记为$r_{ij_i}$，在$A_j$中出现，其中$j_i\in J$。因为不然的话，反例$e_i^-$就会被L所覆盖。因此，由这m个元素组成EM($e^+$)中的一条路。

例如，图3中的路对应于公式$[x_2\neq 1][x_3\neq 1]=[x_2=0, 2][x_3=0, 2]$，再利用图2(a)中的映射，转换成人的特征，正好是2.1节中第1组的描述规则。

1.4　扩张矩阵的应用

示例式学习的目标是寻找在反例集背景下覆盖正例集的规则，而这些规则一般是以公式的析取的形式给出的，这里一个公式是一个选择子或多个选择子的合取式，因此为了使学得的规则尽可能简单精炼，一些最优化问题具有重要意义。其中如下两个最优化问题最为重要：

1．最优覆盖问题（MCV）——生成具有最少数目公式的覆盖。

2．最简公式问题(Mcomp)——生成具有最少数目选择子及变元取值的公式。

作者在[7]文中使用扩张矩阵法将最优覆盖问题和最简公式问题归约到一个NP-难题，从而证明了这两个问题都是NP难题，第一次从理论上表明了示例式学习算法的复杂度，同时还对这两个问题给出了近似算法[7]。

布尔函数的极小化问题是开关理论中的重要问题，它在开关电路和超大规模集成电路设计中起着重要作用，但是已存在的几种布尔函数的表示方法及相应的极小化算法都具有很大局限性，作者使用扩张矩阵方法证明了布尔函数的表格表示法中，最优极小化问题是NP难题，并给出了一个基于扩张矩阵的近似算法[9]。

示例式学习系统GS[5]使用两个启发式策略，即概括和特化策略。所谓概括就是先使用那些覆盖尽可能多正例的属性，特化是排除上述概括过程中，覆盖的反例，这个方法是对AQ11，AQ15，AE1，AE5的综合，其某些方面性能超过AQ11及ID3[6]。把GS和一个概念聚类系统CLUSTER/2[10]联合使用可以做成一个从实例中自动发现规则的系统 Thought/KD1[11]。

2．决策树法及其应用

2.1　决策树

决策树是用于分类处理的树状结构。对于对象集C决策树的构造过程如下：

若C空则我们任意地把它归为正例类或反例

类，若所有E中对象均属于正例类或均属于反例类，则决策树就是只含这一类名的叶子，否则C包含若干正例类成员和若干反例类成员，我们选择一个属性，把C划分成C_1，C_2…，C_n，这里C_i包含我们选择的属性取第i个值所划分C中的成员。如此一直处理下去，最后的输出是一棵树，这棵树的每个叶子带有一个类名，每个中间节点是按所选取的属性可能取值范围的分支点。

为了说明这个过程，我们仍用2.1节中的例子（图1），下面我们用"＋"表示其属于正例类、"－"表示其属于反例类，我们选用属性发色做树根，其决策树为：

（发色）

红色	金色	黑色
（高，红色，蓝色＋）	（矮，金色，蓝色＋） （高，金色，蓝色＋） （矮，金色，灰色＋） （高，金色，黑色－） （矮，金色，黑色－）	（矮，黑色，蓝色－） （高，黑色，蓝色－） （高，黑色，灰色－）

（眼睛）

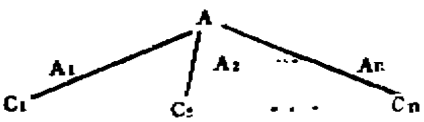
蓝色	灰色	黑色
（矮，金色，蓝色＋） （高，金色，蓝色＋）	（矮，金色，灰色＋）	（高，金色，黑色－） （矮，金色，黑色－）

对于发色取值为黑色和红色，它已归为一类，则不再继续这个过程，对于发色取值为金色的一支，还没有完全分类，再取属性眼睛为子树根，即可把这一支完成分类。现在它的每一个叶结点都属于同一类，我们注意到选用不同的属性做树根或子树根，所构造的决策树复杂性不一样，即将一对象集划分的代价不同，为使形成的决策树代价最小，Quinlan引入信息论中熵的概念选择树根及子树根。决策树可视作一个信息源，对于给定对象，它产生一条对象类的消息。属性的选择所基于的似然假设是决策树的复杂性强依赖于该消息所传送的信息量，若这些消息的概率是P^+和P^-，则该消息的消息量为：

$$M(c) = -P^+\log_2 p^+ - P^-\log_2 P^-$$

对已知对象集C，我们可用相关频率逼近这些概率，于是P^+就是C中正例集的比率，定义M({ }) ＝0。

下面考虑选择A作为树根或子树根的情况，设A的属性值为$A_1 \cdots A_n$且互不相同。



于是所期望信息内容是：

B(C，A)＝Σ属性A的值A_i的概率•M(C_i)

其中用相关频率替代概率，则下一个包含信息最多的属性可选择：

M(C)－B(C，A)取最大值的A

以上面的例子为例，C中有4个对象属于正例类，5个对象属于反例类，

$$M(c) = -4/9 \bullet \log_2(4/9) - 5/9 \bullet \log_2(5/9) = 0.99 \text{ bits}$$

选择发色为树根，则属性值为红色和黑色的已归一类，所以这两分支所获信息量为0。属性值为金色的分支所获信息量为：

$$-2/5 \bullet \log_2(2/5) - 3/5 \bullet \log_2(3/5) = 0.971 \text{ bits}$$

其期望信息内容为：

B(C，"发色")＝5/9•0.971＋1/9•0＋3/9•0＝0.539 bits

该属性所获信息为：0.991－0.539＝0.452bits

选择身材为树根，决策树为：

（身材）

高	矮
（高，红色，黄色：＋）	（矮，金色，蓝色：＋）
（高，金色，蓝色：＋）	（矮，金色，灰色：＋）
（高，金色，黑色：－）	（矮，黑色，蓝色：－）
（高，黑色，蓝色：－）	（矮，金色，蓝色：＋）
（高，黑色，灰色：－）	

"高"的分支所获信息量为：

$$-2/5 \bullet \log_2(2/5) - 3/5 \bullet \log_2(3/5) = 0.971 \text{ bits}$$

"矮"的分支所获信息量为：

$$-2/4 \bullet \log_2(2/4) - 2/4 \bullet \log_2(2/4) = 1 \text{ bits}$$

于是期望信息内容为：B(C"身材")＝5/9×0.971＋4/9×1＝0.984 bits

该属性所获信息为：(0.991)－(0.9840)＝0.007bits

同理可算出选用属性眼睛为树根时所获信息为0.23，故我们选择发色做树根。从它的决策树我们很容易形成描述正例集和反例集的规则。

2.2 决策树的应用

从上面可以看出决策树方法简单，易于实现，算法效率较高，它是目前最好的示例学习算法，用决策树方法实现的示例学习系统ID3[4]是目前最为成功的示例式学习系统，它被广泛地用于知识获取系统中。Quinlan使用这个方法研究了国际象棋残局学习程序，国际象棋权威曾经认为，在国王-车对国王-马的残局中,除少数布局外，大都是平局，但Quinlan的学习程序却发现了还有许多情况是国王-车一方获胜[4]。

作者的示例式学习系统GS[6]，本质上是产生决策树中覆盖正例集尽可能多的那条路，这可使得总的公式数目较少。但GS却丧失了决策树的并行性，即用时间复杂度换取空间复杂度。

3. 结束语

我们讨论了示例式学习的两个主要方法扩张矩阵法和决策树法及其应用。扩张矩阵法具有严谨而完整的理论，是解决示例式学习理论问题的良好工具，在理论上有重要价值，其算法简单，结果以规则形式显式表达，但速度较决策树慢。决策树法简单而易于实现，算法快速而有效，因而广为使用，它是目前最著名的示例式学习方法，但它的结果只能隐式表达，缺乏透明性。

参考文献

〔1〕 R. S. Michalski etal, Machine learning: challenges of the eighties, in Machine learning: An Artificial Intelligence Approach, R. S. Michalski, J. G. Corbonell and T. M. Mitchell (eds), Morgan Kaufmann Publishers Inc 1986.

〔2〕 R. S. Michalski and J. B. Larson, Selection of Most Representative Training Examples and Incremental Generation of VL1 Hypotheses: The Underlying Methodology and the Description of Programs ESEL and AQ11, Report No. 867, Department of Computer Science, University of Illinois, Urbana, May 1978.

〔3〕 Jiarong Hong, I. Mozetic and R. S. Michalski, AQ15: Incremental Learning of Attribute-Based Description from Examples, and User's Guide, ISG 86-5, UIUCDCSF-86-94, Department of Gemputer Science, University of Illinois, Urbana, January 1986.

〔4〕 J. R. Quinlan, Learning Efficient Classification Procedures and Their Application to Chess and Games, In Machine Learning: An Artificial Intelligence Approach, R. S. Michalski, J. G. Carbonell and T. M. Mitchell (Eds.), Palo Alto, CA: Tioga Press, 1983.

〔5〕 D. Haussler, Quantifying Inductive Bias: AI Learning Algorithms and Valiant's Learning Framework, Artificial Intelligence VOL36. No.2, 1988.

〔6〕 J. R. Hong, C. Uhrik: A new attribute-based learning algorithm GS and a comparison with existing algorithms, Journal of Computer Science and Technology Vol.4. No3, 1989.

〔7〕 Jiarong Hong, AE1: An Extension Approximate Method for the General Covering Problem, International Journal of Computer and Information Science, VOL. 14, No. 6, 1985.

〔8〕 洪家荣 示例式学习及多功能学习系统AE5 计算机学报 89.2

〔9〕 洪家荣 SWT——一个基于示例式学习的布尔函数极小化系统，电子学报，VOL. 17 No.6, 1989.

〔10〕 R. S. Michalski and Stepp. R, Automated constructure of Classification: Conceptual Clustering Versus Nemerical Taxoncemy, IEEE VOL. PAM-5 No. 4 July 1983

〔11〕 J. R. Hong, C. J. Mao, Thought/KD1: Theory, Methodology and Implementation of Knowledge discovery. in Proc. of IJCAI-89. Workshop on Knowledge Discovery in Databases Aug. 1989 Detroit. Mi. U. S. A.