

# 二进制可分辨矩阵的最小属性约简算法

李龙澍,王慧萍,徐 怡

(安徽大学 计算机科学与技术学院,安徽 合肥 230039)

**摘 要:**指出传统分辨矩阵的不足,给出了二进制可分辨矩阵的定义以及二进制可分辨矩阵元素集合的形成算法。精简了分辨矩阵元素。在此基础上,提出了一种基于二进制可分辨矩阵的最小属性约简算法。该约简算法以属性频率为选择条件,按照普通可分辨矩阵生成属性约简的原理,但以不同的形式,更少的存储空间,最终可以获得一个最小属性约简。通过对一个汽车数据库的数据进行属性约简,并将结果与其他算法的结果进行比较,证明该算法是可行有效的。

**关键词:**决策表;分辨矩阵;二进制可分辨矩阵;属性约简

**中图分类号:**TP301.6

**文献标识码:**A

**文章编号:**1673-629X(2010)06-0093-04

## Algorithm for the Least Attribute Reduction of Binary Discernibility Matrix

LI Long-shu, WANG Hui-ping, XU Yi

(Department of Computer Science and Technology, Anhui University, Hefei 230039, China)

**Abstract:** In the paper, some shortcomings of traditional discernibility matrix are pointed out. The definitions of the binary discernibility matrix and algorithm of the binary discernibility matrix collection are provided. On this basis, the least attribute reduction algorithm based on the binary discernibility matrix, which uses the attribute frequency as selected information and according to original principle of attribute reduction, is presented by the different forms and less memory space. Finally one of the least attribute reduction can be gained. In the end of the paper, an experimental result of car database, comparing with the result of other algorithm, shows the provided algorithm is feasible.

**Key words:** decision table; discernibility matrix; binary discernibility matrix; attribute reduction

## 0 引 言

粗糙集(Rough Set)理论是由波兰科学家 Z. Pawlak 于 1982 年提出的<sup>[1]</sup>,它具有很强的定性分析能力,是一种能够有效地处理不精确、不连续性和不完整性数据的数学工具<sup>[2]</sup>。粗糙集理论是建立在分类机制的基础上的,它将分类理解为在特定空间上的等价关系,而等价关系构成了对该空间的划分。粗糙集理论将知识理解为对数据划分的结果,每一被划分的集合称为概念。它的主要思想是利用已知的知识库,将不精确或不确定的知识用已知知识库中的知识来近似刻画。粗糙集理论是目前使用较多的一种归纳学习方法,已应用于机器学习、知识发现、数据挖掘、决策支持与分析、专家系统、归纳推理和模式识别等许多科学和

工程领域<sup>[3]</sup>。

属性约简是粗糙集理论的重要内容之一。一般来说,由现实获取的信息形成的知识库,其中的知识并不是同样重要的,存在冗余的信息,这些冗余的部分对规则的有效提取产生很大的影响。因此,有必要对现实中获取的知识进行属性约简。所谓的属性约简,就是在保持知识库分类能力不变的条件下,删除其中不相关或不重要的属性(知识)。通常情况下,一个知识库的约简并不是唯一的。而人们通常真正使用的约简只有一个。约简的关键是求出决策表的最小条件属性集<sup>[4]</sup>。如何从多个属性约简中选出比较优越的约简?这是文中所要解决的问题之一。由于属性个数的多少直接影响了规则的繁简和编码长度,因此要得到最简洁的决策规则首先必须得到一个包含最小属性的约简,即最小约简<sup>[5]</sup>。

文中根据粗糙集理论中的分辨矩阵,在求解所有约简的过程中,利用整数二进制位运算,避开了大量的符号逻辑运算,在未求解出所有约简的情况下,给出了

收稿日期:2009-10-21;修回日期:2010-01-22

基金项目:安徽省自然科学基金(090412054);安徽省重大科技专项(08010201002)

作者简介:李龙澍(1956-),男,教授,博士生导师,研究方向为不精确信息处理和智能软件。

求解最小约简的算法。最终,通过实例说明该算法的可行性。

## 1 粗糙集理论的基本概念

下面主要介绍文中所要用到的粗糙集理论中几个基本的概念。

定义1 四元组  $S = (U, A, V, f)$  是一个知识表达系统,其中  $U$ :对象的非空有限集合,称为论域; $A$ :属性的非空有限集合; $V = \bigcup_{a \in A} V_a$ ,  $V_a$  是属性  $a$  的值域; $f: U \times A \rightarrow V$  是一个信息函数,它为每个对象的每个属性赋予一个信息值,即任意  $A \in a, X \in u, f(x, a) \in V_a$ 。知识表达系统也称为信息系统。通常也用  $S = (U, A)$  来代替  $S = (U, A, V, f)$  [6]。

定义2 设  $S = (U, A, V, f)$  为一知识表达系统,  $A = (C \cup D)$ ,  $C \cap D = \emptyset$ ,  $C$  称为条件属性集,  $D$  称为决策属性集。具有条件属性和决策属性的知识表达系统称为决策表 [6]。

定义3 给定知识表达系统  $S = (U, A, V, f)$ , 称  $B$  ( $B$  是  $A$  的子集) 是  $A$  的一个约简,如果  $B$  满足:

(1)  $\text{ind}(B) = \text{ind}(A)$ ;

(2) 对任意  $K, K$  包含于  $B, \text{ind}(K) \neq \text{ind}(A)$ 。

定义4 [7] 在决策表  $S = (U, C \cup D, V, f)$  中,  $f(x_i, a_j)$  表示元素  $x_i$  在条件属性  $a_j$  上的取值,  $f(x_i, d)$  表示元素  $x_i$  在决策属性  $d$  上的取值。如果满足  $f(x_i, a) = f(x_j, a) \rightarrow f(x_i, d) = f(x_j, d), (a \in C; x_i, x_j \in U; d \in D)$ , 则  $S$  为一致性决策表,否则为非一致决策表。

如果  $U_i$  是一致性的基本块,且  $|U_i| > 1$  (其中  $| \cdot |$  为基数),则基本块存在冗余。若  $S' = (U', C \cup D, V', f')$ , 满足  $U' = \{x'_1, x'_2, \dots, x'_m\}, x'_i \in U_i (i = 1, 2, \dots, m)$ , 则  $S'$  为简化的决策表 [7]。

为了降低分辨矩阵所占用的存储空间,根据定义4,需要对决策表  $S$  进行简化。

定义5 [8] 决策表信息系统  $S = (U, A, V, f)$  中,其中  $A = C \cup D$ , 设  $U/A = \{[x'_1]_A, [x'_2]_A, \dots, [x'_m]_A\}$ , 记  $U' = \{x'_1, x'_2, \dots, x'_m\}$ , 则简化后的决策表信息系统记为:  $S' = (U', C \cup D, V, f)$ 。

定义6 [7] 简化分辨矩阵是定义在简化决策表  $S' = (U', C \cup D, V, f')$  上的一个  $m \times m$  的对称矩阵  $M_{m \times m} = (m_{ij})$ 。矩阵元素定义为:

$$m_{ij} = \begin{cases} \{a \mid a \in C: f(x_i, a) \neq f(x_j, a)\}, D(x_i) \neq D(x_j) \\ \emptyset, & \text{其它情况} \end{cases}$$

这是文献[7]给出的改进的分辨矩阵。这种改进

的分辨矩阵的定义相对于原有的定义更加简单,并且不用计算决策属性对条件属性的正域。

但是这种简化分辨矩阵仍然存在一些问题和不足:该分辨矩阵用属性名称来表示矩阵元素,在实验中消耗了大量的存储空间,并且所生成的分辨矩阵在求解属性约简时,需要大量的符号逻辑运算。另外,分辨矩阵中元素所形成的析取范式,根据符号逻辑运算中的吸收律,元素中一部分是可以被另一部分吸收的,因此分辨矩阵中存在大量重复元素,浪费了存储空间。

## 2 简化二进制可分辨矩阵的形成

针对以上改进分辨矩阵存在的问题,基于简化决策表信息系统  $S'$  定义一种二进制可分辨矩阵。

定义7 在简化决策表  $S' = (U', C \cup D, V, f)$  上定义一个  $m \times m$  的二进制可分辨矩阵  $BM_{m \times m} = (m_{ij})$ , 矩阵中元素定义为一个二进制数,该二进制数的位数为  $|C|$  个:

$$m_{ij} = \begin{cases} \{a \text{ 所对应的二进制位为} \\ 1 \mid a \in C: f(x_i, a) \neq f(x_j, a)\}, D(x_i) \neq D(x_j) \\ \{a \text{ 所对应的二进制位为} \\ 0 \mid a \in C: f(x_i, a) = f(x_j, a)\}, D(x_i) \neq D(x_j) \\ 0, \text{其他情况} \end{cases}$$

该二进制可分辨矩阵将分辨矩阵中的元素用二进制整数表示,其所占存储空间比用符号表示的矩阵要少,另外,在由分辨矩阵形成属性约简时,将符号逻辑运算转换成了位逻辑运算,从而提高了效率。

对于分辨矩阵中存在的重复元素,文中设计了一个算法用于形成比较简化的二进制分辨矩阵元素。

算法1 对于简化决策表  $S' = (U', C \cup D, V, f)$ , 其二进制可分辨矩阵形成算法描述如下:

输入:简化决策表  $S' = (U', C \cup D, V, f)$ ;

输出:二进制可分辨矩阵集合  $BM$ 。

(1) 考察每一对由两个对象组成的有序对  $(x_i, x_j)$ , 根据定义7, 形成二进制数  $m_{ij}$  ( $m_{ij}$  的初值为零)。

(2) 对于每产生一个非零的  $m_{ij}$ , 考虑如何将其添加到集合  $BM$  中,  $BM$  的初值为空集:

1) 如果  $BM$  是空集, 则直接将  $m_{ij}$  添加到  $BM$  中, 即令  $BM \leftarrow BM \cup \{m_{ij}\}$ ; 否则对于  $BM$  中的每一个元素  $m'$ , 如果  $m'$  和  $m_{ij}$  进行二进制与位运算的结果不等于  $m'$  且不等于  $m_{ij}$ , 即  $m'$  不包含于  $m_{ij}$ , 且  $m_{ij}$  也不包含于  $m'$ , 则直接将  $m_{ij}$  添加到  $BM$  中, 即令  $BM \leftarrow BM \cup \{m_{ij}\}$ ;

2) 如果  $m'$  和  $m_{ij}$  进行二进制与位运算的结果等

于  $m_{ij}$ , 即集合  $BM$  中有一个或多个  $m'$  包含  $m_{ij}$ , 则将一个或多个  $m'$  从集合  $BM$  中删除, 最后将  $m_{ij}$  添加到  $BM$  中, 即  $BM \leftarrow (BM - \{m'\}) \cup \{m_{ij}\}$ ;

3) 如果  $m'$  和  $m_{ij}$  进行二进制与位运算的结果等于  $m'$ , 即  $BM$  中的  $m_{ij}$  包含  $m'$ , 则不对  $BM$  做任何改动;

(3) 输出二进制分辨矩阵  $BM$ , 算法结束。

根据 HU<sup>[9]</sup> 给出的可分辨矩阵的求核方法的思想, 可以给出二进制可分辨矩阵的核定义。

定义 8 对于简化决策表  $S' = (U', C \cup D, V, f)$ , 其核定义为:

$BMC_{Core}(C) = \{a_r \mid a_r \in C, m_{ij} \text{ 中属性 } a_r \text{ 对应二进制位为 } 1, \text{ 其余属性二进制位为 } 0\}$

该算法所形成的二进制可分辨矩阵, 其中的每个元素  $m_{ij}$  都有较强的分辨能力, 且不存在重复元素。同时, 该分辨矩阵很容易找到核属性, 即只有一个二进制位是 1 的分辨矩阵元素中的属性。但是, 要付出一定的代价, 那就是在形成分辨矩阵过程中, 消耗了一定的时间复杂度。

例如:  $BM = \emptyset$ , 依顺序加入 01000011(bgh), 11010111(abdfgh), 11001110(abefg), 10001010(aeg), 根据算法 1 求  $BM$ 。

a. 加入 01000011(bgh) 时, 因为  $BM = \emptyset$ , 直接将 01000011(bgh) 加入  $BM$ ,  $BM = \{01000011(bgh)\}$ ;

b. 加入 11010111(abdfgh) 时, 因为  $BM$  中的 01000011(bgh) 包含于 11010111(abdfgh), 所以  $BM$  不改变,  $BM = \{01000011(bgh)\}$ ;

c. 加入 11001110(abefg) 时, 因为  $BM$  中的 01000011(bgh) 和 11001110(abefg) 之间不存在包含关系, 所以将 11001110(abefg) 添加到  $BM$ ,  $BM = \{01000011(bgh), 11001110(abefg)\}$ ;

d. 加入 10001010(aeg) 时, 因为  $BM$  中的 11001110(abefg) 包含 10001010(aeg), 所以将 11001110(abefg) 从  $BM$  中删除, 再将 10001010(aeg) 添加到  $BM$  中,  $BM = \{01000011(bgh), 10001010(aeg)\}$ 。

### 3 基于属性个数的最小属性约简

一般的分辨矩阵形成约简是: 首先, 通过将分辨矩阵中的元素表示成析取式, 再将这些析取式进行合取形成区分函数; 其次, 将区分函数运用吸收律进行化简, 再将合取范式区分函数转换成析取范式, 析取范式中的每个子合取式就是一个约简。

从析取范式中子合取式的形成过程, 我们发现, 每个子合取式的属性是从合取范式中每个子析取式中存

在的属性选一个组合成的, 而子合取式就是约简。一个决策表信息系统的约简可能存在多个。但是通常真正在实际中应用的约简只有一个。由于属性个数的多少直接影响了规则的繁简和编码长度, 因此要得到最简洁的决策规则首先必须得到一个包含最小属性的约简, 即最小约简。

要形成最小属性约简, 从合取范式的子析取式中挑选哪一个属性, 成为要考虑的问题。传统基于可分辨矩阵的属性重要性是由它出现频率决定的, 一个属性出现越频繁, 它的可分辨能力就越强; 可分辨矩阵中元素项越短, 该项的重要性也越大<sup>[10]</sup>。根据该思想, 文中将给出基于属性个数的最小属性约简。

算法 2 基于属性个数的最小属性约简。

输入: 简化决策表  $S' = (U', C \cup D)$ ;

输出: 最小属性约简 Red。

(1) 根据算法 1 产生二进制可分辨矩阵  $BM$ ;

(2) 计算  $C$  中每个属性  $a$  在二进制可分辨矩阵  $BM$  中出现的频率, 即属性  $a$  在所有二进制可分辨矩阵中出现的次数;

(3) 从二进制可分辨矩阵  $BM$  的每个元素  $m$  中, 选择一个属性  $c$  加入 Red 中, 该属性  $c$  在矩阵元素  $m$  中存在, 且是矩阵元素  $m$  中所有元素频率最高的属性, 如果最高频率属性出现多个, 按照数据表中属性的顺序, 从左到右选择; 如果元素  $m$  是核属性, 即只有一个二进制位是 1 的矩阵元素, 就必须选择核属性, 所以核属性一定在约简 Red 中;

(4) 集合 Red 就是所求的约简。

### 4 实例分析

例如: 文献[11]中的汽车数据库有 8 个条件属性。Make - model, cyl, door, display, compress, power, trans, weight, 分别用 a b c d e f g h 表示, 决策属性 mileage 记为 i。

由表 1, 根据经典分辨矩阵的定义, 得经典分辨矩阵  $M$  如表 2 所示。

根据算法 1, 得表 1 的  $BM = \{01000011(bgh), 01101000(bce), 00000111(fgh), 00010010(dg), 00001001(eh), 00011000(de), 10000000(a)\}$ , 再根据算法 2, 计算出每个属性的频率如表 3 所示。得出最小约简为  $Red = \{a, e, g\}$ 。

表 3 属性频率表

属性	e	g	h	b	d	a	c	f
频率	3	3	3	2	2	1	1	1

表 1 汽车数据库表

U	a	b	c	d	e	f	g	h	i
1	USA	6	2	Medium	High	High	Auto	Medium	Medium
2	USA	6	4	Small	Medium	Medium	Manual	Medium	Medium
3	USA	4	2	Medium	High	Medium	Auto	Medium	Medium
4	USA	4	2	Medium	Medium	Medium	Manual	Medium	Medium
5	USA	4	2	Medium	Medium	High	Manual	Medium	Medium
6	USA	6	4	Medium	Medium	High	Auto	Medium	Medium
7	USA	4	2	Medium	Medium	High	Auto	Medium	Medium
8	USA	4	2	Medium	High	High	Manual	Light	High
9	JAPAN	4	2	Small	High	Low	Manual	Light	High
10	JAPAN	4	2	Medium	Medium	Medium	Manual	Medium	High
11	JAPAN	4	2	Small	High	High	Manual	Medium	High
12	JAPAN	4	2	Small	Medium	Low	Manual	Medium	High
13	JAPAN	4	2	Small	High	Medium	Manual	Medium	High
14	USA	4	2	Small	High	Medium	Manual	Medium	High

表 2 经典分辨矩阵

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1														
2	0													
3	0													
4	0	0	0											
5	0	0	0	0										
6	0	0	0	0	0									
7	0	0	0	0	0	0								
8	bgh	bdefh	fgh	efh	eh	bcegh	egh							
9	abdfgh	abcefh	adefgh	adeffh	adeffh	abcdefgh	adefgh	0						
10	abefg	abcd	aeg	a	af	abcfg	afg	0	0					
11	abd	abce	adfg	adef	ade	abcdeg	adeg	0	0	0				
12	abdefg	abcf	adefg	adf	adf	abcdfg	adfg	0	0	0	0			
13	abdfg	abce	adg	ade	adef	abcdefg	adefg	0	0	0	0	0		
14	bdfg	bce	dg	de	def	bcdefg	defg	0	0	0	0	0	0	

首先,从分辨矩阵  $M$  和二进制矩阵  $BM$  中的元素个数,能够看出,  $BM$  所占的存储空间要比  $M$  少,生成  $BM$  的时间复杂度为  $O(|C| \times |U/C|^2 \times |BM|)$ 。其次,文中约简的结果是文献[10]中的约简结果之一,且是属性个数最小的约简,因此,文中的算法是可行的。

## 5 结束语

文中分析了文献[7]中给出的简化分辨矩阵存在的问题和不足,给出了一种二进制可分辨矩阵的定义以及二进制可分辨矩阵形成的算法。该算法是在简化

决策表信息系统  $S'$  基础上进行的,因此,不存在重复对象,为了尽量减少占用存储空间,在产生矩阵元素过程中,根据吸收律,对矩阵元素进行了化简;同时,文中还在二进制可分辨矩阵的基础上提出了一种最小属性约简算法,该算法利用属性出现的频率,选择加入约简集合的属性。由于每个元素都要选择一个属性,所以核属性一定存在于该约简中。而约简中非核属性为出现频率最高的属性。通过实验证明,该算法是可行的。但是,算法仍然存在不足之处,就是算法形成的约简只能说是最小约简,不能保证是最佳约简,这个问题有待进一步研究。

(下转第 100 页)

的问题,并且对 miRNA 基因的从头预测(de novo prediction)很大程度上尚未解决。如何准确定位成熟 miRNA 也尚待研究。且此类方法的一大缺陷是没有一个系统的方法把候选 miRNA 序列的初级结构和二级结构的信息融合起来,用以捕捉序列数据中可能存在的微弱“信号”。

在已有方法中所采用的一些启发式参数也有待进一步研究,如:MiRscan 方法中,候选 miRNA 前体经 RNAfold 折叠后的最小折叠自由能阈值下限设为 25(即  $\Delta G_{\text{folding}} \leq -25$  kcal/mole),但经过对已知的 miRNA 前体进行折叠后发现,许多最小折叠自由能小于该阈值,如线虫 miRNA 前体 cel-mir-261 经折叠后其  $\Delta G_{\text{folding}} = -7.08$  kcal/mole。因此,如何避免设立这些人为参数成为识别算法的一个重要方面。

基于以上各方面的考虑,机器学习方法成为识别算法的一个很好选择。

目前研究的热点在于如何将序列和结构信息综合起来,以达到更好的分类效果。已经提出的基于 string kernel 和 graph kernel 的算法是个很好的尝试。

#### 参考文献:

- [1] Lim L P, Lau N C, Weinstein E G, et al. The microRNAs of *Caenorhabditis elegans* [J]. *Genes Dev.*, 2003, 17: 991 - 1008.
- [2] Lai E C, Tomancak P, Williams R W, et al. Computational identification of *Drosophila* microRNA genes [J]. *Genome Biol.*, 2003(4): 1 - 20.
- [3] Sewer A, Paul M, Landgraf P, et al. Identification of clustered microRNAs using an ab initio prediction method [J]. *Bioinformatics*, 2005(6): 267 - 281.
- [4] Wang X J, Reyes J L, Chua Nam-Hai, et al. Prediction and i-

dentification of *Arabidopsis thaliana* microRNAs and their mRNA targets [J]. *Genome Biol.*, 2004(5): 1 - 15.

- [5] Wang X W, Zhang J, Li F, et al. MicroRNA Identification Based on Sequence and Structure Alignment [J]. *Bioinformatics* 2005, 21(18): 3610 - 3614.
- [6] Xue Chenghai, Li Fei, He Tao, et al. Classification of real and pseudo microRNA precursors using local structure - sequence features and support vector machine [J]. *Bioinformatics* 2005 (6): 310 - 317.
- [7] Yang Liang Huai, Hsu Wynne, Lee Mong Li, et al. Identification of MicroRNA Prediction via SVM [C] // Proceeding of the 4th Asia - Pacific Bioinformatics Conference. Taipei, Taiwan: [s. n.], 2006: 267 - 276.
- [8] Kim Sung - Kyu, Nam Jin - Wu, Rhee Je - Keun, et al. mi-Target: microRNA target gene prediction using a support vector machine [J]. *Bioinformatics*, 2006(7): 411 - 422.
- [9] Leslie C S, Eskin E, Noble W S. The spectrum kernel: a string kernel for SVM protein classification [C] // Proc. Pac. Bio-comput. Symp. [s. l.]: [s. n.], 2002: 1441 - 1448.
- [10] Leslie C S, Eskin E, Cohen A, et al. Mismatch string kernels for discriminative protein classification [J]. *Bioinformatics* 2004, 20(4): 467 - 476.
- [11] Teramoto R, Aoki M, Kimura T, et al. Prediction of siRNA functionality using generalized string kernel and support vector machine [J]. *FEBS Lett.*, 2005, 579(13): 2878 - 2882.
- [12] Yasubumi, Sakakibara. Kernel Functions for RNA sequence analyses [C] // 2nd Taiwan - Japan Bilateral Symposium on Bioinformatics. [s. l.]: [s. n.], 2006.
- [13] Karklin Y, Meraz R F, Holbrook S R. Classification of Non - Coding RNA Using Graph Representations of Secondary Structure [C] // Pacific Symposium on Biocomputing. [s. l.]: [s. n.], 2005.

(上接第 96 页)

#### 参考文献:

- [1] Pawlak Z. Rough Sets [J]. *International Journal of Computer and Information Sciences*, 1982, 11(5): 341 - 356.
- [2] ZHANG Wen-xiu, MI Ju-sheng, WU Wei-zhi. Approaches to knowledge reductions in inconsistent systems [J]. *International Journal of Intelligent Systems*, 2003, 18(9): 989 - 1000.
- [3] 赵荣利, 崔志明, 陈建明. 一种改进的基于差别矩阵的属性约简方法 [J]. *计算机技术与发展*, 2006, 16(11): 32 - 33.
- [4] 汪小燕, 杨思春. 一种基于分辨矩阵的新的属性约简算法 [J]. *计算机技术与发展*, 2008, 18(2): 77 - 78.
- [5] 陈鑫影, 邱占芝. 基于可分辨重要度的属性约简算法 [J]. *大连交通大学学报*, 2008, 29(4): 83 - 84.

- [6] 张文修, 吴伟志, 梁吉业, 等. 粗糙集理论与方法 [M]. 北京: 科学出版社, 2001.
- [7] 田卫东, 周创德, 胡学刚, 等. 基于简化分辨矩阵的粗糙集属性约简算法 [J]. *计算机科学*, 2008, 35(3): 210 - 211.
- [8] 葛浩, 杨传健, 李龙澍. 一种改进的基于二进制可分辨矩阵属性约简算法 [J]. *计算机技术与发展*, 2008, 18(8): 13 - 14.
- [9] HU X H, Cercone N. Learning in relational databases: A rough set approach [J]. *Computational Intelligence*, 1995, 11(2): 323 - 337.
- [10] 胡可云. 基于概念格核粗糙集的数据挖掘方法研究 [D]. 北京: 清华大学, 2001.
- [11] 李佩, 刘玉树. 一种粗糙集属性约简算法 [J]. *计算机工程与应用*, 2002, 38(5): 15 - 19.