# 关于属性约简和集合覆盖问题的探讨

陈彩云　李治国

（南开大学组合数学研究中心，天津 300071）

E-mail：chencaiyun@eyou.com

**摘　要**　论文探讨了粗糙集的属性约简和集合覆盖问题之间的联系。通过构造信息系统的相关矩阵将粗糙集的属性约简问题与集合覆盖问题联系起来，从而将粗糙集的属性约简问题简化为集合覆盖问题。然后用几个定理及其证明说明了这种联系是存在的。基于这种联系，推断出求最小属性约简问题算法的近似度的上下界为 $\ln(|U'|)-\ln\ln(|U'|)+O(1)$ 和 $(1-o(1))\ln(|U'|)$。最后，利用两个范例分别演示了如何具体地构造相关矩阵以及如何将解决集合覆盖问题的思想和方法应用到解决属性约简问题中来，由此推理如果将文献 5 中的解决集合覆盖问题的启发式方法应用到解决最小属性约简中，属性约简的复杂度为 $o(2m^3+m^2)$，并且能以 78%的'概率'得到最小属性约简。

**关键词**　属性约简　集合覆盖　NP-hard 问题　粗糙集

**文章编号** 1002-8331-（2004）02-0044-03　**文献标识码** A　**中图分类号** TP18

# A Study of Reduction of Attributes and Set Covering Problem

Chen Caiyun　Li Zhiguo

（Center for Combinatorics of Nankai University，Tianjin 300071）

**Abstract：** This paper discusses the relationship between the reduction of attributes and set covering by constructing the relation matrix and then gives several theorems to interpret this relationship.The authors deduce the upper bound and the lower bound of approximation is $\ln(|U'|)-\ln\ln(|U'|)+O(1)$ and $(1-o(1))\ln(|U'|)$ respectively.In the end they give two examples to demonstrate how to construct the relation matrix and how the ideas and methods of set covering can be applied into the problem of reduction of attributes.And at the same time we can infer that complexity and the precision of the reduction of attributes is $o(2m^3+m^2)$ and 78 percent respectively from the heuristic algorithm appearing in reference 5.

**Keywords：** Reduction of Attributes，Set Covering，NP-hard problem，Rough sets

## 1　引言

粗糙集是波兰 Z.Pawlak 教授提出的一种数据分析理论[1]。该理论为发现重要数据结构和复杂对象的分类提供了强有力的基础。目前已经在数据挖掘、模式识别、人工智能和分类领域等有很广泛的应用。它的核心内容之一就是属性约简。国内外很多专家对之进行了研究，已经知道，求粗糙集的最小属性约简是一个 NP-hard 问题[1]。尽管粗糙集的理论研究已取得很大的进展，但在属性约简方面的具体实现算法还不多见。已有的，如基于信息熵、基于差别矩阵等算法虽然在某些问题上取得了相当的成效，但到目前为止，还没有一个公认的、高效的属性约简算法[4]。论文将粗糙集属性约简问题和组合优化中的经典问题——集合覆盖问题联系起来，在它们之间找出了一种对应关系，使得求最小属性约简问题转化成求最小集合覆盖问题。集合覆盖问题是经典的 NP-hard 问题，国内外已有很多著名的专家对此进行了深入的研究，理论发展比较成熟。作者想通过此篇论文，将属性约简问题简化为集合覆盖问题，并将解决集合覆盖问题的算法和思想应用到解决属性约简问题中，希望能够进一步地理解粗糙集属性约简的本质。

论文的组织结构：首先介绍了一些粗糙集和集合覆盖问题的基本知识，然后通过几个定理给出了这两者之间的联系，最后举例说明如何具体运用集合覆盖的思想和方法解决属性约简问题，从而进一步阐述了两个问题之间的联系，最后给出了总结及将来要研究的工作。

## 2　粗糙集基本理论[2]

### 2.1　信息系统

信息系统由 4 元集组成，记为 $S=<U,Q,V,f>$，其中：

$U$：由个研究对象 $\{x_1,x_2,\cdots,x_N\}$ 组成的非空集合，称为闭域；

$Q$：由 $n$ 个属性 $\{q_1,q_2,\cdots,q_n\}$ 组成的有限非空集合；

$V=\cup_{q\in Q}V_q$：表示 $Q$ 中所有属性的值域，其中 $V_q$ 是属性 $q\in Q$ 的值域。

$f$：$U\times Q\rightarrow V$：全决策函数，使得对于任一 $x\in U,q\in Q$，有 $f(x,q)\in V_q$。通过 $f$ 作用，信息系统 $S$ 能用一个有限的数据表表示，表的第 $i$ 行研究对象 $x_i$ 和第 $j$ 列属性 $q_j$ 有对应的值。

### 2.2　等价类

首先介绍一下不可区分关系的定义，对信息系统 $S=<U,Q,V,f>$ 和 $Q$ 中的任意子集 $A\subseteq Q$，定义不可区分关系：

$$IND(A)=\{(x,y)\in U\times U: f(x,q)=f(y,q),for\ all\ q\in A\}$$

也就是说，任给 $(x,y)\in IND(A)$，$x$ 和 $y$ 具有相同的属性值。这样根据不可区分关系 $IND(A)$，将 $U$ 中属性值相同的对

基金项目：国家自然科学基金资助；国家 973 重点基础研究规划项目（编号：G19980306）资助；教育部、科技部基金资助

作者简介：陈彩云（1975-），女，南开大学组合数学研究中心 2001 级博士生，研究方向：组合数学与数据挖掘。李治国（1977-），男，南开大学组合数学研究中心 2000 级硕士生，研究方向：组合数学与应用。

象归为同一等价类，于是闭域 $U$ 被分成 $r$ 个不同的等价类 $X_1$，$X_2$，$\cdots X_r$。对属性集 $A$ 中每个属性 $q_j$ $(j=1,2,\cdots,n)$ 而言，每个等价类中的任两个对象 $x$、$y$，他们对应的属性值都相等即：$\forall x$，$y \in X_i$，有 $f(x,q_j)=f(y,q_j)$。由这样不同的等价类组成的等价类族 $X=\{X_1,X_2,\cdots X_r\}$ 也称为闭域 $U$ 关于属性集 $A$ 的一个分类。在文中，不妨取 $A=Q$，即 $X=\{X_1,X_2,\cdots X_r\}$ 是关于所有属性的一个分类。

### 2.3 属性约简和核

众所周知，信息系统中的属性并不是同等重要的，甚至其中某些知识是冗余的。特别是，当信息系统中的数据是随机采集时，其冗余性更为普遍。所谓属性约简，就是在保持信息系统分类能力不变的条件下，删除其中的冗余属性。一般地讲，一个信息系统的属性约简不是唯一的。人们期望找到具有最少属性的约简，即最小约简。它在文中，实际上就是从 $Q$ 中找出个数最少的属性集，使得这些属性能区分闭域 $U$ 上的 $r$ 个等价类。

所有约简的交集称为粗糙集的核。由此知道核这个概念的用处有两方面：首先它可以作为所有约简的计算基础，因为核包含在所有的约简之中，并且计算可以直接进行；其次可解释为在属性约简时它是不能消去的属性特征集合。

上述的粗糙集理论中所有的概念和运算都是通过代数学的等价关系和集合运算来定义的，被称之为粗糙集理论的代数表示。在代数表示下，粗糙集理论的很多概念与运算的直观性较差，人们不容易理解其本质。另外在这种表示下，目前还没有属性约简的高效算法[4]。同时由于直观性较差，目前已经提出的很多属性约简的方法，几乎都没有对算法的近似度进行分析。而集合覆盖问题是经典的组合优化问题，已经有很多比较成熟有效的算法。用集合覆盖的方法来分析解决属性约简问题，比较直观明了，而且容易分析算法的近似程度。在下面一节中将简单介绍一下集合覆盖问题。

## 3 集合覆盖问题

首先简单描述一下最小集合覆盖问题的定义。

定义 1：最小集合覆盖问题：$S$ 是一个集合，$S_1$，$S_2$，$\cdots S_m$ 是 $S$ 的子集，且构成 $S$ 的覆盖，即 $\bigcup_{i=1}^{m} S_i=S$，求最小的覆盖。

从最小集合覆盖问题的定义中可以看到，最小集合覆盖的前提是所有的子集能够把 $S$ 覆盖。所以，在这样的前提下，最小集合覆盖是一定存在的。反之，如果所有的子集都不能够覆盖 $S$，那显然也就不存在最小的集合覆盖了。

集合覆盖问题是经典的 NP-hard 问题之一[3]，很多科学家对之已经做出了深刻的研究，其理论发展已经比较成熟，解决此问题的算法很多，比如 SCHF 算法、HVC 算法等等[5]。Petr Slavík[6]证明该问题的近似度可以达到 $\ln(|S|)-\ln\ln(|S|)+O(1)$。但是 Uriel Feige[7]，在 1998 年证明了集合覆盖问题不能被近似度为 $(1-o(1))\ln(|S|)$ 或更低的近似算法近似。由此可以看到，集合覆盖问题的研究已经很成熟了。如果属性约简能被化简成集合覆盖问题，那么解决和分析集合覆盖问题的思想方法都可以用到解决粗糙集的属性约简上面来。作者将在下一节阐述属性约简与集合覆盖的联系。

## 4 属性约简和覆盖问题

在介绍属性约简和集合覆盖问题的联系之前，首先给出一个信息系统的相关矩阵的定义。

定义 2：信息系统 $S=<U,Q,V,f>$ 的相关矩阵 $M$

假定信息系统 $S=<U,Q,V,f>$ 的等价类族 $X=\{X_1,X_2,\cdots X_r\}$ 和属性集 $Q=\{q_1,q_2,\cdots q_n\}$，由等价类族 $X$ 中任意两个不同等价类组成集合记为：$U'=\{(X_i,X_j)|X_i,X_j \in X, i<j \le r\}$，显然 $|U'|=\frac{r(r-1)}{2}$。定义一个 $\frac{r(r-1)}{2}$ 行 $n$ 列的矩阵 $M$，其中行代表 $U'$ 中的元素，列表示 $Q$ 中的元素。矩阵第 $i$ 行第 $j$ 列元素 $m_{ij}$ 取值如下表示：

任给 $u_i=(X',X'') \in U'$，$q_j \in Q$，有：

$$m_{ij}=\begin{cases}1 & f(x,q_j) \neq f(y,q_j), \forall x \in X', y \in X'' \\ 0 & f(x,q_j)=f(y,q_j), \forall x \in X', y \in X''\end{cases}$$

由这样的 $m_{ij}$ 组成的矩阵 $M=(m_{ij})$，称为信息系统 $S$ 的相关矩阵。

信息系统的相关矩阵具有定理 1 和定理 2 所述的两个性质。

定理 1 相关矩阵 $M$ 中没有全零的行向量。

证明：采用反证法。如果相关矩阵的第 $i$ 行的所有元素均为 0，即 $m_{ij}=0$，$(j=1,2,\cdots,n)$，根据相关矩阵的定义知道 $U'$ 中元素 $u_i=(X',X'') \in U'$ 对应的两个不同的等价类 $X'$、$X''$，他们中的所有对象对于 $Q$ 中的每一个属性都有相同的取值，即 $\forall x \in X'$，$y \in X''$，对于 $Q$ 中的每一个属性 $q_j$，都有 $f(x,q_j)=f(y,q_j)$。根据等价类的定义知道 $X'$ 和 $X''$ 是相同的一类，但这与 $X' \neq X''$ 产生矛盾，从而相关矩阵 $M$ 没有全零的行向量。

定理 2：在相关矩阵 $M$ 中，如果 $m_{ij}=1$，则属性 $q_j$ 可以区分开 $u_i=(X',X'') \in U'$ 对应的两个等价类 $X'$ 和 $X''$，否则 $q_j$ 不能区分 $X'$ 和 $X''$。

证明：根据矩阵 $M=(m_{ij})$ 的定义可知：如果 $m_{ij}=1$，对 $U'$ 中第 $i$ 行元素 $u_i=(X',X'')$，$\forall x \in X'$，$y \in X''$，有 $f(x,q_j) \neq f(y,q_j)$，这说明属性 $q_j$ 能将两个等价类 $X'$，$X''$ 区分开。否则属性 $q_j$ 不能将这两个等价类区分开。结论成立，证毕。

为了方便叙述，如果属性 $q_j$ 能将 $u_i=(X',X'') \in U'$ 对应的两个等价类 $X'$ 和 $X''$ 区分开，那就说属性 $q_j$ 可以区分 $u_i$ 或者说 $u_i$ 可以被 $q_j$ 区分，否则就说属性 $q_j$ 不能区分 $u_i$ 或者说 $u_i$ 不能被 $q_j$ 区分。

由定理 1 和定理 2 可以直接得到推论 1。

推论 1：对于 $U'$ 中任一个元素 $u=(X',X'') \in U'$，至少存在一个属性 $q$ 将 $u$ 区分。

再给出一个属性的区分集合的定义。

定义 3：区分集合

由 $U'$ 中所有能被 $q_j$ 区分的 $u=(X',X'') \in U'$ 的全体组成的集合叫作属性的区分集合，记为 $S_{q_j}$。可以表示为：$S_{q_j}=\bigcup_{i=1}^{|U'|}\{u_i|m_{ij}=1, u_i=(X',X'') \in U'\}$。

有了区分集合的定义，就可以把每个区分集合看作是 $U'$ 的一个子集合，事实上所有的区分集合都能够覆盖集合 $U'$，如定理 3 所示。

定理 3：所有属性的区分集合构成集合 $U'$ 的覆盖，即 $U'=\bigcup_{j=1}^{n} S_{q_j}$。

证明：用反证法来证明。假设所有的区分集合不能覆盖集合 $U'$，则说明 $U'$ 中存在一个元素 $u=(X_i,X_j) \in U'$ 不属于 $Q$ 中任

何一个属性的区分集合。再结合区分集合的定义及定理 1 和定理 2，可以推出 $u$ 不能被 $Q$ 中任何属性区分，这个结论与推论 1 矛盾，所以结论成立。证毕。

定理 3 说明了区分集合与全集 $U'$ 满足最小覆盖的前提条件，所以最小覆盖是存在的。从而，有以下的定理。

定理 4：$U'$ 的一个集合覆盖等价于粗糙集的一个属性约简。

证明：由区分集合的定义可以知道，任何一个 $U'$ 的覆盖都可以把 $U'$ 中的所有元素区分开，即可以把 X 中的等价类两两区分开。再根据属性约简的定义，这些区分集合所对应的属性集就是粗糙集的一个属性约简。反之依然成立。证毕。

定理 4 说明了 $U'$ 的集合覆盖与粗糙集的属性约简的对应关系。任意一个属性约简都对应于 $U'$ 的一个集合覆盖，任意一个 $U'$ 的集合覆盖都对应一个粗糙集的属性约简，所以粗糙集的最小属性约简就等价于集合 $U'$ 的最小集合覆盖。又因为最小覆盖问题是 NP-hard 问题，所以也可以从另外的一个角度说明计算最小属性约简的复杂性。于是从定理 4，直接得到以下两个推论。

推论 2：粗糙集的最小属性约简 $U'$ 对应的最小集合覆盖，同时 $U'$ 的最小集合覆盖也对应粗糙集的最小属性约简。

推论 3：粗糙集的最小属性约简问题是 NP-hard 问题。

本节讨论了属性约简与集合覆盖问题的联系，证明了求粗糙集最小属性约简等价于求的最小覆盖。第 3 节中提到了 Petr Slavík 和 Uriel Feige 分别给出了求最小集合覆盖问题算法的近似度的上下界。那么可以推断出相应的粗糙集最小属性约简问题的算法的近似度的上下界为 $\ln(|U'|)-\ln\ln(|U'|)+O(1)$ 和 $(1-o(1))\ln(|U'|)$。

## 5 示例

本节通过两个范例分别来演示如何从原始的信息系统构造其相关矩阵，以及如何将解决集合覆盖问题的思想应用到解决属性约简问题上面来。

例 1 将构造参考文献 2 中提到的一个信息系统的相关矩阵 $M$ 和区分集合。原始信息系统 $U$ 表示如表 1。

表 1

| $U$ | $P$ | $Q$ | $r$ | $s$ |
|---|---|---|---|---|
| $x_1$ | 0 | 1 | 1 | 0 |
| $x_2$ | 2 | 1 | 0 | 1 |
| $x_3$ | 0 | 0 | 0 | 1 |
| $x_4$ | 1 | 0 | 2 | 0 |
| $x_5$ | 0 | 1 | 1 | 0 |
| $x_6$ | 2 | 1 | 0 | 1 |
| $x_7$ | 0 | 1 | 1 | 0 |
| $x_8$ | 1 | 0 | 2 | 1 |
| $x_9$ | 0 | 0 | 0 | 1 |

该系统构成的 5 个等价类分别为：$X_1=\{x_1,x_5,x_7\}$，$X_2=\{x_2,x_6\}$，$X_3=\{x_3,x_9\}$，$X_4=\{x_4\}$，$X_5=\{x_8\}$，根据定义 2 得知相关矩阵 $M$ 为：

根据定义 3 给出属性的区分集合分别为：

属性 p 的区分集合 $S_p=\{u_1,u_3,u_4,u_5,u_6,u_7,u_8,u_9\}$。

属性 q 的区分集合 $S_q=\{u_2,u_3,u_4,u_5,u_6,u_7\}$。

属性 r 的区分集合 $S_r=\{u_1,u_2,u_3,u_4,u_6,u_7,u_8,u_9\}$。

属性 s 的区分集合 $S_s=\{u_1,u_2,u_4,u_6,u_8,u_{10}\}$。

很明显可以看到 $M$ 还具有上面定理所述的几个性质：

(1) $M$ 中没有全零的行向量；

(2) 所有属性的区分集全可以覆盖 $U'$；

(3) $S_p$ 和 $S_s$ 是 $U'$ 的覆盖，所以属性 p 和 s 是原信息系统的属性约简。

表 2 信息系统的区分矩阵 $M$

| $U'$ | $p$ | $q$ | $r$ | $s$ |
|---|---|---|---|---|
| $u_1=\{X_1,X_2\}$ | 1 | 0 | 1 | 1 |
| $u_2=\{X_1,X_3\}$ | 0 | 1 | 1 | 1 |
| $u_3=\{X_1,X_4\}$ | 1 | 1 | 1 | 0 |
| $u_4=\{X_1,X_5\}$ | 1 | 1 | 1 | 1 |
| $u_5=\{X_2,X_3\}$ | 1 | 1 | 0 | 0 |
| $u_6=\{X_2,X_4\}$ | 1 | 1 | 1 | 1 |
| $u_7=\{X_2,X_5\}$ | 1 | 1 | 1 | 0 |
| $u_8=\{X_3,X_4\}$ | 1 | 0 | 1 | 1 |
| $u_9=\{X_3,X_5\}$ | 1 | 0 | 1 | 0 |
| $u_{10}=\{X_4,X_5\}$ | 0 | 0 | 0 | 1 |

例 2：参考文献[5]提到了一些在构造集合覆盖问题的算法时常用的完备策略，这里就对这些完备策略，给出其在粗糙集属性约简中的一些解释。

通俗地讲，一个完备的策略是，当问题的实例满足策略条件时，把该实例化简为更易解的实例，并且新实例的最优化解都是原实例的最优化解，即不增加最优化解的策略[5]。关于完备策略的严格定义请参见文献[5]。

文献[5]中提到的三个在解决集合覆盖问题用到的完备策略为：

完备策略 1 如果 $S_1,S_2,\cdots S_m$ 中的一个集合 $S_i=S$，则选择 $S_i$ 作为最优化覆盖中的唯一一个集合 $S_i$。

完备策略 2 如果存在 $x\in S$，$x$ 只属于 $S_1,S_2,\cdots S_m$ 中的一个集合 $S_i$，则选择 $S_i$ 作为最优化覆盖中的一个集合 $S_i$。

完备策略 3 如果 $S_i\subseteq S_j$，则排除 $S_i$。

下面说明这三种完备策略也可以作为构造属性约简的启发式算法时的三个完备策略。

对于完备策略 1，该策略说明在属性约简中如果存在一个属性能将 $U$ 中所有的等价类区分开，则很显然，该属性就是粗糙集的最小属性约简。

对于完备策略 2，在粗糙集中，如果 $U$ 中的两个等价类 $X_i$ 和 $X_j$，他们只能被某个属性 $q$ 区分，那么 $q$ 一定在粗糙集的最小属性约简中。由前面的定义知道，所有这样的属性构成粗糙集的核。这符合粗糙集理论中这样一个思想：计算属性约简一般从计算他的核开始。

策略 3 也很明显，在属性约简中，比如属性 $q_1$ 只能够区分 $(X_1,X_2)$，而属性 $q_2$ 不仅仅可以区分 $(X_1,X_2)$，也可以区分 $(X_1,X_3)$，那 $q_1$ 可以马上被排除掉，因为 $q_2$ 可以代替 $q_1$ 来区分 $(X_1,X_2)$。特别地，如果任意两个等价类都不能被属性 $q_1$ 区分，则 $q_1$ 可以马上被排除。这样的情形在相关矩阵中的表现为，属性 $q_1$ 对应的区分集合为空集，即 $q_1$ 的列向量是全零的。

通过上述三个策略的对应，可以将文献 5 中提到的启发式算法 (SCHF) 应用到属性约简问题中来。文献 5 通过构造一个从 $S_1,S_2,\cdots S_m$ (即文中的区分集合) 到实数轴的一个映射 $F$ 作为启发函数来计算最小覆盖，算法详见[5]。该算法的复杂度为 $O(m^3+m^2)$，并且文献 5 通过实验得到该方法能够以 78% 的 '概率' 得到最优解，从而可以推出 SCHF 也可以以同样的 '概率' 得到最小的属性约简。

由此可见，解决集合覆盖问题的一些方法和策略是可以用

(下转 84 页)

对 $\forall x' \in S^{\alpha}(x)$，有 $x \in S^{\alpha}(x')$，

$x \in S^{\alpha}(x') \cap X \neq \emptyset$，即 $x' \in \overline{R}^{\alpha}(X) \subseteq Y$，

即 $x' \in Y$，由 $x'$ 之任意性知 $S^{\alpha}(x) \subseteq Y$，即

$x' \in \underline{R}^{\alpha}(Y)$ 故 $x \subseteq \underline{R}^{\alpha}(Y)$

从而 $x \subseteq \underline{R}^{\alpha}(Y) \Leftrightarrow \overline{R}^{\alpha}(X) \subseteq Y$

由性质 5 知：$X$ 包含于 $Y$ 的下近似当且仅当 $X$ 的上近似包含于 $Y$。

## 5 实例

作者给出一个不完备信息系统的实际例子，并计算给定 $X$ 的上、下近似集。

设不完备信息系统（$U,AT$）如表，有 $U=\{1,2,3,4,5,6,7,8,9,10\}$，$AT=\{A,B,C,D,E\}$，任取 $a \in AT, x \in U$，有 $a(x)=0$ 或 $a(x)=1$ 或 $a(x)=*$。

取 $\alpha=0.8$ 有

$S^{\alpha}(1)=\{1,3,7,10\}$　　$S^{\alpha}(2)=\{2,4,5,6,8\}$

$S^{\alpha}(3)=\{1,3,7,10\}$　　$S^{\alpha}(4)=\{2,4,6\}$

$S^{\alpha}(5)=\{2,5,6,8\}$　　$S^{\alpha}(6)=\{2,4,5,6\}$

$S^{\alpha}(7)=\{1,3,7,8,9\}$　　$S^{\alpha}(8)=\{2,5,7,8\}$

$S^{\alpha}(9)=\{7,9\}$　　$S^{\alpha}(10)=\{1,3,10\}$

设 $X=\{2,4,5,6,8\}$，则

$\underline{R}^{\alpha}(X)=\{2,4,5,6\}$

$\overline{R}^{\alpha}(X)=\{2,4,5,6,7,8\}$

$\underline{R}^{\alpha}[\underline{R}^{\alpha}(X)]=\{4,6\}$

$\overline{R}^{\alpha}[\underline{R}^{\alpha}(X)]=\{2,4,5,6,8\}$

$\underline{R}^{\alpha}[\overline{R}^{\alpha}(X)]=\{2,4,5,6,8\}$

$\overline{R}^{\alpha}[\overline{R}^{\alpha}(X)]=\{1,2,3,4,5,6,7,8,9\}$

$POS(X)_{\alpha}=\underline{R}^{\alpha}(X)=\{2,4,5,6\}$

$NEG(X)_{\alpha}=U-\overline{R}^{\alpha}(X)=\{1,3,9,10\}$

$BND(X)_{\alpha}=\overline{R}^{\alpha}(X)-\underline{R}^{\alpha}(X)=\{7,8\}$

$\rho_R(X)_{\alpha}=1-4/6=1/3$

显然有性质 4，即：

$\underline{R}^{\alpha}[\underline{R}^{\alpha}(X)] \subseteq \underline{R}^{\alpha}(X) \subseteq \overline{R}^{\alpha}[\underline{R}^{\alpha}(X)] \subseteq X$

$\subseteq \underline{R}^{\alpha}[\overline{R}^{\alpha}(X)] \subseteq \overline{R}^{\alpha}(X) \subseteq \overline{R}^{\alpha}[\overline{R}^{\alpha}(X)]$

成立

表 1

| U \ AT | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | 0 | 1 | 1 | 0 | 1 |
| 2 | 1 | 0 | 0 | * | 0 |
| 3 | 0 | 1 | * | 0 | 1 |
| 4 | * | 0 | 0 | 1 | 1 |
| 5 | 1 | 0 | 1 | 0 | * |
| 6 | 1 | * | * | 1 | 1 |
| 7 | 0 | 1 | 1 | * | 0 |
| 8 | 1 | * | 1 | 0 | 0 |
| 9 | 0 | 1 | 0 | 1 | 0 |
| 10 | 0 | 1 | 0 | 0 | 1 |

## 6 结论

论文在文[1]将集对分析方法用于刻画不完备信息系统的基础上，定义了另一种不完备信息系统的上、下近似运算并得到了比文[1]更完善的上、下近似运算的一些基本性质；通过一个简单的例子说明了用上述方法的可行性，在粗糙集用于研究不完备信息系统研究方面做了一定推广。当不完备信息中对象太少或空值过多时，这种方法仍是适用的；同时，$\underline{R}^{\alpha}(X)$ 与 $\overline{R}^{\alpha}(X)$ 中 $\alpha$ 的选取是主观的，它取决于人们对同一程度的不同要求。$\alpha$ 越大，对立度越低；$\alpha$ 越小，对立度越高。当然，作者只是将粗集理论与集对思想的结合进行了初步探讨，他们在实际应用中的价值还有待于进一步的深入研究。

（收稿日期：2003 年 2 月）

## 参考文献


1.黄兵，周献中.基于集对分析的不完备信息系统粗糙集模型[J].计算机科学，2002；29（专刊）：1~3

2.张文修等.粗糙集理论与方法[M].北京：科学出版社，2001

3.赵克勤.集对分析其初步应用[M].浙江科学出版社，2000

4.Kryszkiewicz M，Rybinski H.Incomplete aspects in rough set approach[C].In：Proc Of JCIS'98.Raleigh USA，1998：371~374

5.刘清.Rough 集及 Rough 推理[M].北京：科学出版社，2001

6.赵卫东，吴明赞.不完全信息下的粗集拓展[J].计算机科学，2002；29（专刊）：131~134

7.乔斌，郑洪涛等.针对信息不完备性的粗糙集分层递阶约简[J].电路与系统学报，2001；6（2）：78~82


（上接 46 页）

到最小属性约简问题中来的。

## 6 结束语及未来工作

在文中作者对用集合覆盖问题的思想和方法来研究粗糙集的属性约简问题进行了初步探讨。通过构造信息系统的相关矩阵，将粗糙集的属性约简问题简化成为集合覆盖问题，从而使得众多的解决集合覆盖的方法和思想用来解决属性约简问题成为可能。

论文对粗糙集属性约简和集合覆盖之间联系的探讨，现在仅仅局限在理论的研究上，只是一个起步。对于怎样才能高效地构造粗糙集的相关矩阵以及探讨相关矩阵的其它性质，包括集合覆盖的一些具体算法怎样才能高效率地用到属性约简上，并且易于机器实现等等，这些工作都有待进一步去研究。

（收稿日期：2003 年 5 月）

## 参考文献


1.Pawlak Z.Rough Sets[J].International Journal of Computer Sciences，1982；（1）：341~356

2.Krzysztof J Cios，Witold Pedrycz，Roman W Swiniarski.Data Mining Methods for Knowledge Discovery[M].Kluwer Academic Publishers，1998

3.C H Papadimitriou，K Steiglitz.Combinatorial Optimization Algorithms and Complexity[M].1982

4.苗夺谦，王珏.粗糙集理论中概念与运算的信息表示[M].软件学报，1999；（2）：113~116

5.权光日等.集合覆盖问题的启发式算法[J].软件学报，1998；9（2）

6.Petr Slavík.A tight analysis of the greedy algorithm for set cover[J].Journal of Algorithms 1997-11；25

7.Uriel Feige.A threshold of lnn for approximating set cover[J].Journal of the ACM，1998-07；45