昆明理工大学第九届大学生数学建模竞赛题目

（请先阅读 “论文格式要求”）

**A题：关系数据库的概念与规则发现问题**

在形式概念分析（国内也称概念格）理论中，有一种特殊的关系数据库，称为形式背景，它通常记为，其中是对象集，是属性集，为每个对象在各个属性下指定一个取值. 比如，表1给出了一个形式背景，对象集是4个病人，属性集为“：发烧”，“：咳嗽”，“：头痛”和“：呼吸困难”，为每个病人在各个属性下指定了取值“0”或“1”，其中“0”代表病人在该属性下正常，“1”代表病人在该属性下不正常.

**表1** 一个形式背景

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 病人 | ：发烧 | ：咳嗽 | ：头痛 | ：呼吸困难 |
| 1 | 1 | 1 | 0 | 0 |
| 2 | 1 | 1 | 0 | 1 |
| 3 | 0 | 0 | 1 | 1 |
| 4 | 1 | 0 | 1 | 1 |

对于一个形式背景，人们通常关心“概念”发现问题. 所谓“概念”，是指具有明确“外延”和“内涵”的序对，其中和满足：（1）对于中的所有对象，如果它们在某个属性下的取值均为1，那么这个属性就被筛选出来，且所有通过这种方式筛选出来的属性组成的集合恰好为；（2）对于中的所有属性，如果某个对象在它们下的取值均为1，那么这个对象就被筛选出来，且所有通过这种方式筛选出来的对象组成的集合恰好为. 比如，表1中，容易验证序对就是一个概念.

现实中，形式背景还携带有决策属性，称这种配备决策属性的形式背景为决策形式背景，通常记为，它可看作由两个形式背景和组合而成. 比如，表2给出了一个决策形式背景. 相比较于表1，它多了两个决策属性和（），为每个病人在各个决策属性下指定了取值“0”或“1”，其中“0”表示未患该病，“1”表示患有该病.

**表2** 一个决策形式背景

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 病人 | ：发烧 | ：咳嗽 | ：头痛 | ：呼吸困难 | ：患病A | ：患病B |
| 1 | 1 | 1 | 0 | 0 | 1 | 0 |
| 2 | 1 | 1 | 0 | 1 | 1 | 0 |
| 3 | 0 | 0 | 1 | 1 | 0 | 1 |
| 4 | 1 | 0 | 1 | 1 | 0 | 1 |

对于一个决策形式背景，人们通常关心“规则”发现问题. 规则是指，的一个概念与的一个概念形成的内涵对应关系，其中外延和满足. 人们通过经验发现，一个决策形式背景的规则数往往都比较庞大，因此进一步提出了“非冗余规则”. 非冗余规则，是指它满足（1）是规则，（2）不存在的一个概念使得，（3）不存在的一个概念使得. 比如，表2中，就是一条非冗余规则，因为概念和满足上述条件（1），（2）和（3）.

根据上述信息，回答如下问题.

**问题1**. 建立获取一个形式背景所有“概念”的数学模型（方法），并通过实验说明方法的计算效率.

**问题2**. 在问题1的基础上，进一步建立获取一个决策形式背景所有“非冗余规则”的数模模型（方法），并通过实验说明方法的计算效率.

**注**：实验所需的（决策）形式背景可参考<http://archive.ics.uci.edu/ml/datasets.html>上列出的关系数据库，比如Zoo, Balance Scale, Wine, Car Evaluation 等. 温馨提醒：该网站上的关系数据库不能直接使用，因为这些原始数据取值并非为0或1，需利用特定方法先将它们变成取值为0或1的情形.