國立交通大學

資訊管理研究所

博士論文

從最佳化觀點推導多評準分類規則 — 以生物及醫療資訊為例

Induction of Multiple Criteria Classification Rules from Optimization Perspectives — Applied in Biology and Medicine Informatics

研究生: 陳明賢

指導教授: 黎漢林 博士

中華民國九十七年六月

從最佳化觀點推導多評準分類規則 — 以生物及醫療資訊為例

Induction of Multiple Criteria Classification Rules from Optimization Perspectives — Applied in Biology and Medicine Informatics

研究生: 陳明賢 Student: Ming-Hsien Chen

指導教授: 黎漢林 Advisor: Han-Lin Li

國立交通大學

資訊管理研究所

博士論文

A Dissertation

Submitted to Institute of Information Management

College of Management

National Chiao Tung University

in Partial Fulfillment of the Requirements

of the Degree of

Doctor of Philosophy in Information Management

June 2008

Hsinchu, Taiwan, Republic of China

中華民國九十七年六月

從最佳化觀點推導多評準分類規則 — 以生物及醫療資訊為例

學生: 陳明賢　　　　　　　　　　　　　　　　指導教授: 黎漢林

國立交通大學資訊管理研究所博士班

摘　　要

從資料中推導出關鍵的分類規則, 是科學研究的重要任務之一。一條有用的分類規則, 除其是最適外, 應同時滿足三項評準: 高正確度、高支持度、高精簡度。然而, 目前的分類方法, 諸如約略集合理論、類神經網路、分類樹等, 都只能推導得可行解規則, 而非最適規則。此外, 目前的方法推導得的規則只能同時滿足前述三項評準之一。本研究提出一個多評準的模式, 用以在較好的正確度、支持度及精簡度下, 推導得最適分類規則, 其是透過混合 0-1 線性多目標規化模型以推導分類規則。並以一些實際的生物及醫療資料進行測試, 其結果顯示所提方法能比目前方法推導得較佳的分類規則。

關鍵字: 分類規則

Induction of Multiple Criteria Classification Rules from Optimization Perspectives — Applied in Biology and Medicine Informatics

Student: Ming-Hsien Chen Advisor: Han-Lin Li

Institute of Information Management

National Chiao Tung University

Abstract

To induce critical classification rules from observed data is a major task in biological and medical research. A classification rule is considered to be useful if it is optimal and simultaneously satisfies three criteria: is highly accurate, has a high rate of support, and is highly compact. However, existing classification methods, such as rough set theory, neural networks, ID3, etc., may only induce feasible rules instead of optimal rules. In addition, the rules found by existing methods may only satisfy one of the three criteria. This study proposes a multi-criteria model to induce optimal classification rules with better rates of accuracy, support and compactness. A linear multi-objective programming model for inducing classification rules is formulated. Two practical data sets, one of HSV patients results and another of European barn swallows, are tested. The results illustrate that the proposed method can induce better rules than existing methods.

Keywords: classification rules

Acknowledgement

正如當年完成碩士論文時的誌謝, 首先依然是要感謝我的父母、陳建仁先生及葉惠美女士的支持與鼓勵, 並提供一個無後顧之憂的環境, 讓我再次於離開學校多年後, 二度拋下工作, 繼續進修, 全心全意投入本研究。因爲有他們對我的愛護, 我才有可能完成學業。

接著, 要感謝黎漢林教授多年來的悉心指導, 使我不論是在學業、思維或爲人處事上, 皆有長足的進步, 於此特別表示由衷的謝意。

同時, 要感謝曾國雄教授、溫于平教授、陳茂生教授、林妙聰教授及李永銘教授的指正與建議, 及林信雄教授的英文寫作指導, 使本論文得以更加完善。

此外, 要感謝陳煇煌教授、蘇宜芬老師及柯宇謙老師鼎力關照, 介紹兼課機會, 使我在博士修業期間, 不但生計不虞匱乏, 且能娶妻生子、並完成學業。

再者, 要感謝運籌管理研究室的伙伴們, 昶瑞、宇謙、麗菁、浩鈞、芸珊、嘉輝、曜輝、婉瑜等人, 陪我一同走過這漫長的研究之路, 無論是生活上的扶持, 或是學業上的切磋砥厲, 都給我莫大的助益。

最後, 我要將接下來的篇幅, 留給我的妻子、李益芬, 因爲有她的支持、體諒及無怨的付出, 我才能熬過低潮, 完成這個研究。結婚這幾年來, 聚少離多, 她不但經常孤單在家, 還歷經患甲亢、懷孕、生子、育嬰、再懷孕, 然她都不畏艱苦、一肩扛起, 讓我全心投入學業。今日的成果, 全因有她成就, 在此, 我不但要感謝她, 還要將這論文獻給她。

Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Research Background

The induction of classification rules[1] from a database has been one of the major issues in the biological and medical research domains. Given a data set with several objects, where each object has some attributes and belongs to a specific class, the induction of rules is to find a combination of attributes which can well describe the features of a specific class. There are three criteria for evaluating the quality of a rule.

(i) **Accuracy**. A good rule which fits a specific class had better not cover objects of other classes.

(ii) **Support**. A good rule which fits a specific class should be supported by most objects of such a class.

(iii) **Compactness**. A good rule should be expressed in a compact way. That means that the less the number of attributes used, the better the rule is.

Currently, there are some well-known methods for classification, especially the rough-set-based method and the decision-tree-based method.

In Hvidsten et al. (2003), rough sets were used as the theoretical foundation for its methodology to learn the rule-based biological process from gene expression time profiles. It reported a systematically supervised learning approach to predict a biological process from the time series of gene expression

[1]Instead of using the term "classification rule", the term "rule" is used for short in the following of this article.

data and biological knowledge. Biological knowledge is expressed using gene ontology and this knowledge is associated with discriminatory-expression-based features to form minimal decision rules. In Beynon and Buchanan (2003), which used variable precision rough sets, a variant of rough sets, to do the gender classification of the European barn swallow. Slowinski (1992), Tsumoto (1999), Li and Wang (2004), Tay and Shen (2002), Shen and Loh (2004), etc., also used the rough-set- based method to get rules.

In Geurts et al. (2005), the decision-tree-based method was used for proteomic mass spectra classification. They proposed a systematic approach based on decision-tree-ensemble methods, which is used to automatically determine proteomic biomarkers and predictive models. Aja-Fernandez et al. (2004)proposed a fuzzy ID3 decision-tree methodology by which the natural language descriptions of the TW3 method for bone age assessment is translated into an automatic classifier. And in Zhang et al. (2001), the decision-tree-based method was used for classifying normal or tumor tissues, etc.

Both the rough set based method and the decision tree based method are heuristic algorithms, which are computationally effective in inducing rules. However, there are two shortcomings for these two methods:

(i) They may find only some feasible rules, instead of inducing optimal rules.

(ii) For most cases, they may find only rules satisfying a single criterion such a more accuracy rate or a more support rate, instead of inducing rules to satisfy multiple criteria.

1.2 Review of Some Existing Methods

There are many well-known methods for classification. Two methods are reviewd here.

1.2.1 Review of Rough Set Theory

Rough set theory (RST) proposed by Pawlak (1982) is a methodology for rules discovery in the database. It operates on an information system[2] which is made up of objects for which certain characteristics (i.e., condition attributes[3]) are known. Objects with the same condition attribute values are classified into equivalence classes or condition classes. The objects are each grouped into a particular category with respect to the decision attribute[4] value. Those classified into the same category are in the same decision class. The rule discovery process in RST involves simplifying the decision tables with the elimination of superfluous attributes and values of attributes, and finding out simple rules related to the condition and decision attributes. When an object is classified using the rules discovered, it is assumed to be a correct classification. A variant of RST, variable precision rough sets (VPRS), which incorporates probabilistic decision rules, has been developed by Ziarko (1993). It has been applied in various fields to induce rules.

[2]The meaning of the term "information system" in RST is synonymous with the term "data set" in this study.

[3]The meaning of the term "condition attribute" in RST is synonymous with the term "attribute" in this study.

[4]The meaning of the term "decision attribute" in RST is synonymous with the term "class index" in this study.

1.2.2 Review of ID3

The ID3 proposed by Quinlan (1986) is a popular decision tree method of inducing rules. It is based on the greedy algorithm of entropy reduction in constructing the decision tree. Attributes leading to substantial entropy reduction (or information gain) are included as condition attributes[5] to partition the data. A condition attribute of the largest amount of entropy reduction is placed closer to the root and is used for the next level partitioning. Sometimes filters may be set up so that only attributes with information gain greater than a certain threshold will be selected in constructing the decision tree. Variants of ID3 include C4.5 and C5 Quinlan (1993), which treat both discrete and continuous variables.

1.3 Research Objectives

Using mathematical programming approaches to solve classification problems are current trends. Sun and Xiong (2003) proposed a mathematical programming approach for gene selection and tissue classification; however, it focused on two classes of classification and could not guarantee to obtain globally optimal solutions. Li and Fu (2005) developed a linear programming technique to solve DNA consensus sequence identification problems by finding an optimum consensus sequence. It was computationally more efficient and guaranteed to reach the global optimum.

This study proposes a multi-criteria model to induce optimal rules with better rates of accuracy, support, and compactness. A mixed 0-1 linear multi-objective programming model for inducing rules is formulated. Two

[5]The meaning of the term "condition attribute" in ID3 is synonymous with the term "attribute" in this study, too.

practical data sets, one of HSV (Highly Selective Vagotomy) patient results and the other of European barn swallows, are tested. The results refeal that the proposed method can induce better rules than can current methods.

1.4 Structure of the Dissertation

Chapter 2 reviews some existing methods. Chapter 3 formally formulates the problem this study deals with and introduces the presentation of data and rules in this study. Chapter 4 developes essential propositions and a method to induce rules. It also illustrates the proposed method with some examples. Chapter 5 compares the proposed method with rought-set-based methods and decision-tree-based methods by two practical data sets, the HSV (Highly Selective Vagotomy) patients data set and the European barn swallow data set. Chater 6 introduces a prototype system, which implements the proposed method. The last Chapter makes some discussions and remarks of the study.

Chapter 2 Problem Formulation and Notations

This chapter gives a formal formulation of induction of rules and makes a presentation of data and rules. It also introduces the notations used in this study.

2.1 Problem Formulation

There are n objects $\{x_1, x_2, \ldots, x_n\}$, each of which is characterized by m attributes $\{a_1, a_2, \ldots, a_m\}$ and a class index c. Each attribute has its own domain of values. The p'th value of an attribute a_j is denoted as $a_{j,p}$, which is called the p'th sub-attribute of the attribute a_j in this study. For a specific class, there may exist some rules for it . The l'th rule for a class k is denoted as $R^{k,l}$. A rule may just use some attributes. A rule $R^{k,l}$ is a combination of binary variables $d_{j,p}^{k,l}$ and each $d_{j,p}^{k,l}$ decides whether sub-attribute $a_{j,p}$ is used by $R^{k,l}$ or not. The purpose of this study is to find rules for each class.

2.2 Presentation of Data and Rules

Here we use an example to illustrate the way of presenting data and rules in this study.Consider a data set in Table 2.1 which has five objects $\{x_1, x_2, x_3, x_4, x_5\}$, four attributes $\{a_1, a_2, a_3, a_4\}$, and one class index c. The domains of values of a_1, a_2, a_3, and a_4 are $\{1, 2, 3\}$, $\{1, 2\}$, $\{1, 2, 3, 4\}$, and $\{1, 2, 3\}$, respectively. The domain of values of c is $\{1, 2, 3\}$. In most cases, the attributes in a data set usually consist of a mixture of qualitative and quantitative ones. In this study, all attributes are transformed into ordered

Table 2.1: A small data set

	a_1	a_2	a_3	a_4	c
x_1	2	1	2	3	1
x_2	3	2	1	1	2
x_3	1	2	2	3	2
x_4	1	1	4	2	3
x_5	1	1	3	1	3

qualitative values. To induce rules for each class, first we convert Table 2.1 into another form presented by binary values in Table 2.2 , where $a_{j,p}$ is called the p'th sub-attribute of j'th attribute. An object x_i in Table 2.1 can then be written as:

$$x_i = (a^i_{1,1}, a^i_{1,2}, a^i_{1,3}; a^i_{2,1}, a^i_{2,2}; a^i_{3,1}, a^i_{3,2}, a^i_{3,3}, a^i_{3,4}; a^i_{4,1}, a^i_{4,2}, a^i_{4,3}; c_i),$$

where $a^i_{j,p}$ is 1 if a^i_j (the value of a_p of x_i) equals p; otherwise, $a^i_{j,p}$ is 0. For instance, x_1 is expressed as

$$x_1 = (0, 1, 0; 1, 0; 0, 1, 0, 0; 0, 0, 1; 1).$$

Notation 2.1. For a data set, which is characterized by m attributes. A general form for expressing an object x_i is written as:

$$x_i = (a^i_{1,1}, \ldots, a^i_{1,q_1}; a^i_{2,1}, \ldots, a^i_{2,q_2}; \ldots; a^i_{m,1}, \ldots, a^i_{m,q_m}; c_i), \qquad (2.1)$$

where q_m is the number of sub-attributes of the attribute a_m[6], c_i is the class index of x_i, and $a^i_{j,p}$ are sub-attribute values of x_i, which are binary

[6]For convenience, the total number of sub-attributes of all attributes is denoted as q, i.e., $q = \sum_j q_j$.

Table 2.2: Binary presentation for the data set in Table 2.1

	a_1			a_2		a_3				a_4			c
	$a_{1,1}$	$a_{1,2}$	$a_{1,3}$	$a_{2,1}$	$a_{2,2}$	$a_{3,1}$	$a_{3,2}$	$a_{3,3}$	$a_{3,4}$	$a_{4,1}$	$a_{4,2}$	$a_{4,3}$	
x_1		1		1			1					1	1
x_2			1		1	1				1			2
x_3	1				1		1					1	2
x_4	1			1					1		1		3
x_5	1			1				1		1			3

All blank cells are 0

values. An $a^i_{j,p}$ is 1 if $a^i_j = p$; otherwise, $a^i_{j,p}$ is 0. Clearly, for each object x_i, $\sum_p a^i_{j,p} = 1$ for all j. □

Notation 2.2. A general form of expressing a rule $R^{k,l}$, which is called the l'th rule for the class k, is expressed as:

$$R^{k,l} = (d^{k,l}_{1,1}, \ldots, d^{k,l}_{1,q_1}; d^{k,l}_{2,1}, \ldots, d^{k,l}_{2,q_2}; \ldots; d^{k,l}_{m,1}, \ldots, d^{k,l}_{m,q_m}), \tag{2.2}$$

where $d^{k,l}_{j,p}$ are binary variables specified as: if $a_{j,p}$ is an active sub-attribute for $R^{k,l}$, then $d^{k,l}_{j,p} = 1$; otherwise, $d^{k,l}_{j,p} = 0$. □

Such a binary expression is useful in inducing rules with conjunctive and disjunctive forms.

Definition 2.1. (Support and Non-Violation) Given objects x_i, x_r, and a rule $R^{k,l}$ as represented in equations (2.1) and (2.2), respectively.

(i) Object x_i belongs to a class k (i.e., $c_i = k$): x_i is called "**supporting**" $R^{k,l}$, and $R^{k,l}$ is called "**supported by**" x_i, if $\sum_p a^i_{j,p} d^{k,l}_{j,p} = 1$ for all active attribute a_j.

Table 2.3: Some rules for Table 2.1

	$d_{1,1}$	$d_{1,2}$	$d_{1,3}$	$d_{2,1}$	$d_{2,2}$	$d_{3,1}$	$d_{3,2}$	$d_{3,3}$	$d_{3,4}$	$d_{4,1}$	$d_{4,2}$	$d_{4,3}$
$R^{1,1}$		1		1								
$R^{1,2}$				1								1
$R^{2,1}$					1							
$R^{2,2}$						1						
$R^{3,1}$								1				
$R^{3,2}$									1			
$R^{3,3}$								1	1			

All blank cells are 0

(ii) Object x_r does not belong to a class k (i.e., $c_r \neq k$): x_r is called "**non-violating**" $R^{k,l}$, if $\sum_p a^r_{j,p} d^{k,l}_{j,p} = 0$ for any active attribute a_j. $\square$

Consider the example given in Table 2.1, Table 2.3 is a list of seven rules induced from Table 2.2. (The method of inducing rules are described in next chapter). $R^{1,1}$ is expressed by a binary vector as

$$R^{1,1} = (0,1,0;1,0;0,0,0,0;0,0,0).$$

Table 2.4 is the explanation of these rules. For instance, $R^{1,1}$ means

"*if* $(a_1 = 2)$ *and* $(a_2 = 1)$*, then the objects belong to class* 1".

The ignored attributes in $R^{1,1}$ are a_3 and a_4; the active attributes are a_1 and a_2 (in fact, the active sub-attributes are $a_{1,2}$ and $a_{2,1}$). This rule is supported by object x_1, since for these two active attributes, we have

$$\sum_{p=1\ldots 3} a^1_{1,p} d^{1,1}_{1,p} = a^1_{1,1} d^{1,1}_{1,1} + a^1_{1,2} d^{1,1}_{1,2} + a^1_{1,3} d^{1,1}_{1,3} = 0 \times 0 + 1 \times 1 + 0 \times 0 = 1$$

and

$$\sum_{p=1...2} a^1_{2,p} d^{1,1}_{2,p} = a^1_{2,1} d^{1,1}_{2,1} + a^1_{2,2} d^{1,1}_{2,2} = 1 \times 1 + 0 \times 0 = 1.$$

And it is not violated by object x_2, since for the active attribute a_1, we have

$$\sum_{p=1...3} a^2_{1,p} d^{1,1}_{1,p} = a^2_{1,1} d^{1,1}_{1,1} + a^2_{1,2} d^{1,1}_{1,2} + a^2_{1,3} d^{1,1}_{1,3} = 0 \times 0 + 0 \times 1 + 1 \times 0 = 0.$$

It is also not violated by objects x_3 ,x_4 and x_5. Furthermore, two rules may be integrated into a more general rule. For instance, $R^{3,1}$ and $R^{3,2}$ can be combined as $R^{3,3}$, expressed as

$$\begin{aligned} R^{3,1} \vee R^{3,2} &= (0,0,0;0,0;0,0,1,0;0,0,0) \vee (0,0,0;0,0;0,0,0,1;0,0,0) \\ &= (0,0,0;0,0;0,0,1,1;0,0,0,) \\ &= R^{3,3}. \end{aligned}$$

$R^{3,3}$ means

“*if* $(a_3 = 3$ *or* $4)$, *then the objects belong to class* 3”.

The meaning of the last three columns in Table 2.4 is explained in next chapter.

2.3 Notations and Variables Summary

Here is a summary of the notations and variables adopted in this chapter.

- x_i: the object i in a data set.
- a_j: the attribute j of objects.
- $a_{j,p}$: the p'th sub-attribute of a_j.
- c: the class index of objects.
- c_i: the class index of x_i.

Table 2.4: The explanation for rules in Table 2.3

rule	meaning	active sub-attributes	ignored attributes	supporting objects	non-violated objects	accuracy rate (AR)	support rate (SR)	compactness rate (CR)
$R^{1,1}$	if $(a_1 = 2) \wedge (a_2 = 1)$, then $c = 1$	$a_{1,2}, a_{2,1}$	a_3, a_4	x_1	x_2, x_3, x_4, x_5	1	1	0.73
$R^{1,2}$	if $(a_2 = 1) \wedge (a_4 = 3)$, then $c = 1$	$a_{2,1}, a_{4,3}$	a_1, a_3	x_1	x_2, x_3, x_4, x_5	1	1	0.73
$R^{2,1}$	if $(a_2 = 2)$, then $c = 2$	$a_{2,2}$	a_1, a_3, a_4	x_2, x_3	x_1, x_4, x_5	1	1	1
$R^{2,2}$	if $(a_3 = 1)$, then $c = 2$	$a_{3,1}$	a_1, a_2, a_4	x_2	x_1, x_4, x_5	1	0.5	1
$R^{3,1}$	if $(a_3 = 3)$, then $c = 3$	$a_{3,3}$	a_1, a_2, a_4	x_5	x_1, x_2, x_3	1	0.5	1
$R^{3,2}$	if $(a_3 = 4)$, then $c = 3$	$a_{3,4}$	a_1, a_2, a_4	x_4	x_1, x_2, x_3	1	0.5	1
$R^{3,3}$	if $(a_3 = 3 \vee 4)$, then $g = 3$	$a_{3,3}, a_{3,4}$	a_1, a_2, a_4	x_4, x_5	x_1, x_2, x_3	1	1	0.98

where "$\wedge$" means logic "AND" and "$\vee$" means logic "OR"

- n: the total number of objects in a data set.
- m: the total number of attributes in a data set.
- q: the total number of sub-attributes of all attributes.
- q_m: the number of sub-attributes of a_m.
- a_j^i: the value of a_j of x_i.
- $a_{j,p}^i$: the value of $a_{j,p}$ of x_i.
- $R^{k,l}$: the l'th rule for the class k.
- $d_{j,p}^{k,l}$: a binary variable specified as: if $a_{j,p}$ is an active sub-attribute for the rule $R^{k,l}$, then $d_{j,p}^{k,l} = 1$; otherwise, $d_{j,p}^{k,l} = 0$.

Chapter 3 Proposed Classification Methods

This chapter developes some essential propositions and a method of inducing rules such as those in Table 2.3. It also illustrates the proposed method with some examples.

3.1 Propositions

First, consider the following propositions:

Proposition 3.1. For objects x_i (such $c_i = k$), x_r (such $c_r \neq k$), and a rule $R^{k,l}$ as represented in equations (2.1) and (2.2), respectively, $h^{k,l}$ is denoted the number of ignored attributes by $R^{k,l}$.

(i) $R^{k,l}$ is supported by x_i , if $\sum_j \sum_p a^i_{j,p} d^{k,l}_{j,p} = m - h^{k,l}$.

(ii) $R^{k,l}$ is not violated by x_r, if $\sum_j \sum_p a^r_{j,p} d^{k,l}_{j,p} \leq m - h^{k,l} - 1$.

Proof. From Definition 2.1, we have

(i) $\sum_p a^i_{j,p} d^{k,l}_{j,p} = 1$ for all active attributes a_j while $R^{k,l}$ is supported by x_i. So it is clear that $\sum_j \sum_p a^i_{j,p} d^{k,l}_{j,p} = m - h^{k,l}$.

(ii) $\sum_p a^r_{j,p} d^{k,l}_{j,p} = 0$ for any active attribute a_j while $R^{k,l}$ is not violated by x_r. So it is clear that $\sum_j \sum_p a^r_{j,p} d^{k,l}_{j,p} \leq m - h^{k,l} - 1$.

The proposition is then proven. □

Take x_1 and x_4 in Table 2.1 and $R^{1,1}$ in Table 2.3 for instance. Since

$c_1 = 1$ and

$$\sum_j \sum_p a^i_{j,p} d^{k,l}_{j,p} = (a^1_{1,1}d^{1,1}_{1,1} + a^1_{1,2}d^{1,1}_{1,2} + a^1_{1,3}d^{1,1}_{1,3}) + (a^1_{2,1}d^{1,1}_{2,1} + a^1_{2,2}d^{1,1}_{2,2})$$

$$= (0 \times 0 + 1 \times 1 + 0 \times 0) + (1 \times 1 + 0 \times 0)$$

$$= 2$$

$$= m - h^{k,l},$$

$R^{1,1}$ is supported by x_1. Since $c_4 = 3 \neq 1$ and

$$\sum_j \sum_p a^r_{j,p} d^{k,l}_{j,p} = (a^4_{1,1}d^{1,1}_{1,1} + a^4_{1,2}d^{1,1}_{1,2} + a^4_{1,3}d^{1,1}_{1,3}) + (a^4_{2,1}d^{1,1}_{2,1} + a^4_{2,2}d^{1,1}_{2,2})$$

$$= (1 \times 0 + 0 \times 1 + 0 \times 0) + (1 \times 1 + 0 \times 0)$$

$$= 1$$

$$\leq m - h^{k,l} - 1,$$

x_4 does not violate $R^{1,1}$.

Proposition 3.2. Parameter $h^{k,l}$ is specified as $h^{k,l} = \sum_j \lambda^{k,l}_j$, where $\lambda^{k,l}_j \in \{0,1\}$. $\lambda^{k,l}_j = 1$ if attribute a_j is ignored by a rule $R^{k,l}$; otherwise, $\lambda^{k,l}_j = 0$. The relationships between $d^{k,l}_{j,p}$ and $\lambda^{k,l}_j$ are expressed as the following inequalities:

$$d^{k,l}_{j,p} \leq 1 - \lambda^{k,l}_j, \forall j, p, \tag{3.1}$$

$$1 - \lambda^{k,l}_j \leq \sum_p d^{k,l}_{j,p}, \forall j, \tag{3.2}$$

$$\lambda^{k,l}_j \in \{0,1\}. \tag{3.3}$$

Proof.

- If attribute a_j is ignored by $R^{k,l}$, then $d^{k,l}_{j,p} = 0$ for all p, $\sum_p d^{k,l}_{j,p} = 0$ and $\lambda^{k,l}_j = 1$.

- If attribute a_j is not ignored by $R^{k,l}$, then at least one $d_{j,p}^{k,l} = 1$, $\sum_p d_{j,p}^{k,l} \geq 1$ and $\lambda_j^{k,l} = 0$.

The proposition is then proven. $\square$

Remark 3.1. For objects x_i (such $c_i = k$), x_r (such $c_r \neq k$), and a rule $R^{k,l}$, here we introduce binary variables $u_i^{k,l}$ and $v_r^{k,l}$:

(i) $u_i^{k,l} = 1$, if x_i supports $R^{k,l}$; otherwise $u_i^{k,l} = 0$.

(ii) $v_r^{k,l} = 1$, if x_r does not violate $R^{k,l}$; otherwise $v_r^{k,l} = 0$. $\square$

Proposition 3.3. Let M be a big positive number. For objects x_i (such $c_i = k$), x_r (such $c_r \neq k$), and a rule $R^{k,l}$, there exit $u_i^{k,l}$ and $v_r^{k,l} \in \{0,1\}$ which satisfy the following inequalities:

$$\begin{aligned} M(u_i^{k,l} - 1) + m - h^{k,l} \leq & \sum_j \sum_p a_{j,p}^i d_{j,p}^{k,l} \\ & \leq m - h^{k,l} + M(1 - u_i^{k,l}), \forall i \text{ where } c_i = k, \end{aligned} \tag{3.4}$$

$$\sum_j \sum_p a_{j,p}^r d_{j,p}^{k,l} \leq m - h^{k,l} - 1 + M(1 - v_r^{k,l}), \forall r \text{ where } c_r \neq k. \tag{3.5}$$

Proof.

- If $u_i^{k,l} = 1$, then equation (3.4) is equivalent to Case 1 of Proposition 3.1.
- If $v_r^{k,l} = 1$, then equation (3.5) is equivalent to Case 2 of Proposition 3.1.

The proposition is then proven. $\square$

Consider a data set of n objects. Denote the number of objects belonging to a specific class k as n^k. The definitions of the accuracy rate, support rate and compactness rate are specified as the following.

Definition 3.1. (Accuracy Rate) The accuracy rate of a rule $R^{k,l}$ is specified as

$$AR^{k,l} = \frac{1}{n - n^k} \sum_{r \text{ where } c_r \neq k} v_r^{k,l}.$$

□

It means that if none of object x_r (such $c_r \neq k$) violates the rule (i.e., all $v_r^{k,l} = 1$), then the accuracy rate of the rule is 1. The binary parameter $v_r^{k,l}$ is specified in Remark 3.1.

Definition 3.2. (Support Rate) The support rate of a rule $R^{k,l}$ is specified as

$$SR^{k,l} = \frac{1}{n^k} \sum_{i \text{ where } c_i = k} u_i^{k,l}.$$

□

If all objects x_i (such $c_i = k$) support the rule (i.e., all $u_i^{k,l} = 1$), then its support rate is 1. The binary parameter $u_i^{k,l}$ is specified in Remark 3.1.

Definition 3.3. (Compactness Rate) The compactness rate of a rule $R^{k,l}$ is specified as

$$CR^{k,l} = \frac{1}{m} \left(h^{k,l} + 1 - \frac{\sum_j \sum_p d_{j,p}^{k,l} - 1}{q} \right).$$

□

It implies that if the most compact rule is the rule with only one active sub-attribute (i.e., such $\sum_j \sum_p d_{j,p}^{k,l} = 1$), then $CR^{k,l} = 1$. If different rules have the same numbers of active sub-attributes, then the rule with larger ignored attributes number h is considered more compact than others. By the definition given here, the CR of a rule with larger h will be higher than others.The parameter $h^{k,l}$ is specified in Proposition 3.1.

Remark 3.2.

(i) $0 \leq AR^{k,l} \leq 1$

(ii) $0 \leq SR^{k,l} \leq 1$

(iii) $0 \leq CR^{k,l} \leq 1$ $\square$

The related AR, SR, and CR values for the example rules in Table 2.3 are listed in the last three columns of Table 2.4 . The CR value for $R^{1,1}$ is computed as $\frac{1}{4}(2+1-\frac{2-1}{12}) = 0.73$. $R^{2,1}$ is better than $R^{2,2}$ since it has a higher SR value. $R^{3,3}$ has higher SR than do $R^{3,1}$ and $R^{3,2}$; however, $R^{3,3}$ is not as compact as $R^{3,1}$ and $R^{3,2}$.

3.2 Notations and Variables Summary

Here is a summary of the notations and variables adopted in this chapter.

- $h^{k,l}$: the number of ignored attributes by $R^{k,l}$.
- $\lambda_j^{k,l}$: a binary variable specified as: if a_j is ignored by $R^{k,l}$, then $\lambda_j^{k,l} = 1$; otherwise, $\lambda_j^{k,l} = 0$.
- $u_i^{k,l}$: a binary variable specified as: if x_i supports $R^{k,l}$, then $u_i^{k,l} = 1$; otherwise, $u_i^{k,l} = 0$.
- $v_r^{k,l}$: a binary variable specified as: if x_r dose not violate $R^{k,l}$, then $v_r^{k,l} = 1$; otherwise, $u_i^{k,l} = 0$.
- M: a big positive number.
- n^k: the number of objects which belong to the class k.
- $AR^{k,l}$: the accuracy rate of $R^{k,l}$.

- $SR^{k,l}$: the support rate of $R^{k,l}$.

- $CR^{k,l}$: the compactness rate of $R^{k,l}$.

3.3 Models for Inducing Rules

The program to induce a rule $R^{k,l}$ is formulated as the following linear multiobjective program:

$$\text{Max } AR^{k,l}$$

$$\text{Max } SR^{k,l}$$

$$\text{Max } CR^{k,l}$$

Subject to:

$$\begin{aligned} M(u_i^{k,l}-1)+m-h^{k,l} &\le \sum_j \sum_p a_{j,p}^i d_{j,p}^{k,l} \\ &\le m-h^{k,l}+M(1-u_i^{k,l}), \forall i \text{ where } c_i = k, \end{aligned} \tag{C1}$$

$$\sum_j \sum_p a_{j,p}^r d_{j,p}^{k,l} \le m-h^{k,l}-1+M(1-v_r^{k,l}), \forall r \text{ where } c_r \ne k, \tag{C2}$$

$$d_{j,p}^{k,l} \le 1-\lambda_j^{k,l}, 1 \le j \le m, 1 \le p \le q_j, \tag{C3}$$

$$1-\lambda_j^{k,l} \le \sum_p d_{j,p}^{k,l}, 1 \le j \le m, \tag{C4}$$

$$h^{k,l} = \sum_j \lambda_j^{k,l}, \tag{C5}$$

$$d_{j,p'}^{k,l} + d_{j,p'+2}^{k,l} - 1 \le d_{j,p'+1}^{k,l}, 1 \le p' \le q_j - 2, \tag{C6}$$

$$u_i^{k,l}, v_r^{k,l}, d_{j,p}^{k,l}, \lambda_j^{k,l} \in \{0,1\}, 1 \le i,j \le n, 1 \le j \le m, 1 \le p \le q_j. \tag{C7}$$

The objective of this program is to maximize AR, SR and CR simultaneously, where constrainsts (C1) and (C2) come from Proposition 3.3, and constraints (C3)–(C5) come from Proposition 3.2. The purpose of constraint (C6) is to avoid the discontinuity of active sub-attributes for the same attribute. $AR^{k,l}$, $SR^{k,l}$, and $CR^{k,l}$ are specified in Definitions 3.1, 3.2, and 3.3, respectively. This is a multiple criteria decision problem. There are three typical models for solving this multiobjective program:

(i) A *constraint model*, where two of the three objectives with lower bounds are assigned.

(ii) An *aspiration model*, where aspiration levels are set for the three objectives.

(iii) A *weighting model*, where relative weights are assigned to the three objectives.

All these three models are reformulated below.

Model 3.1. (Specifying the lower bounds for AR and SR)

$$\text{Max } obj = CR^{k,l}$$

Subject to: constraints (C1)–(C7), and

$$AR^{k,l} \geq \overline{AR},$$

$$SR^{k,l} \geq \overline{SR},$$

where $\overline{AR}$ and $\overline{SR}$ are constant, representing the lower bounds of $AR^{k,l}$ and $SR^{k,l}$. □

Example 3.1. Take Table 2.3 for instance. The program to induce a rule $R^{1,l}$ for class 1 using Model 3.1 is formulated below:

$$\text{Max } CR^{1,l} = \frac{1}{4}\left(h^{1,l} + 1 - \frac{\sum_j \sum_p d_{j,p}^{1,l}}{12}\right)$$

subject to:

$$5(u_1^{1,l}-1)+4-h^{1,l} \le d_{1,2}^{1,l}+d_{2,1}^{1,l}+d_{3,2}^{1,l}+d_{4,3}^{1,l} \le 4-h^{1,l}+5(1-u_1^{1,l}),$$

$$d_{1,3}^{1,l}+d_{2,2}^{1,l}+d_{3,1}^{1,l}+d_{4,1}^{1,l} \le 4-h^{1,l}-1+5(1-v_2^{1,l}),$$

$$d_{1,1}^{1,l}+d_{2,2}^{1,l}+d_{3,2}^{1,l}+d_{4,3}^{1,l} \le 4-h^{1,l}-1+5(1-v_3^{1,l}),$$

$$d_{1,1}^{1,l}+d_{2,1}^{1,l}+d_{3,4}^{1,l}+d_{4,2}^{1,l} \le 4-h^{1,l}-1+5(1-v_4^{1,l}),$$

$$d_{1,1}^{1,l}+d_{2,1}^{1,l}+d_{3,3}^{1,l}+d_{4,1}^{1,l} \le 4-h^{1,l}-1+5(1-v_5^{1,l}),$$

$$d_{1,p}^{1,l} \le 1-\lambda_1^{1,l}, \text{ for } p=1,2,3, \tag{3.6}$$

$$d_{2,p}^{1,l} \le 1-\lambda_2^{1,l}, \text{ for } p=1,2, \tag{3.7}$$

$$d_{3,p}^{1,l} \le 1-\lambda_3^{1,l}, \text{ for } p=1,2,3,4, \tag{3.8}$$

$$d_{4,p}^{1,l} \le 1-\lambda_4^{1,l}, \text{ for } p=1,2,3, \tag{3.9}$$

$$1-\lambda_1^{1,l} \le d_{1,1}^{1,l}+d_{1,2}^{1,l}+d_{1,3}^{1,l}, \tag{3.10}$$

$$1-\lambda_2^{1,l} \le d_{2,1}^{1,l}+d_{2,2}^{1,l}, \tag{3.11}$$

$$1-\lambda_3^{1,l} \le d_{3,1}^{1,l}+d_{3,2}^{1,l}+d_{3,3}^{1,l}+d_{3,4}^{1,l}, \tag{3.12}$$

$$1-\lambda_4^{1,l} \le d_{4,1}^{1,l}+d_{4,2}^{1,l}+d_{4,3}^{1,l}, \tag{3.13}$$

$$h^{1,l} = \lambda_1^{1,l}+\lambda_2^{1,l}+\lambda_3^{1,l}+\lambda_4^{1,l}, \tag{3.14}$$

$$d_{1,1}^{1,l}+d_{1,3}^{1,l}-1 \le d_{1,2}^{1,l}, \tag{3.15}$$

$$d_{3,1}^{1,l}+d_{3,3}^{1,l}-1 \le d_{3,2}^{1,l}, \tag{3.16}$$

$$d_{3,2}^{1,l}+d_{3,4}^{1,l}-1 \le d_{3,3}^{1,l}, \tag{3.17}$$

$$d_{4,1}^{1,l}+d_{4,3}^{1,l}-1 \le d_{4,2}^{1,l}, \tag{3.18}$$

$$AR^{1,l} = \frac{1}{5-1}(v_2^{1,l}+v_3^{1,l}+v_4^{1,l}+v_5^{1,l}) \ge \overline{AR},$$

$$SR^{1,l} = \frac{1}{1} u_1^{1,l} \geq \overline{SR},$$

$$u_1^{1,l}, v_2^{1,l}, v_3^{1,l}, v_4^{1,l}, v_5^{1,l}, d_{1,1}^{1,l}, d_{1,2}^{1,l}, d_{1,3}^{1,l}, d_{2,1}^{1,l}, d_{2,2}^{1,l}, d_{3,1}^{1,l}, d_{3,2}^{1,l}, d_{3,3}^{1,l}, d_{3,4}^{1,l},$$
$$d_{4,1}^{1,l}, d_{4,2}^{1,l}, d_{4,3}^{1,l}, \lambda_1^{1,l}, \lambda_2^{1,l}, \lambda_3^{1,l}, \lambda_4^{1,l} \in \{0,1\}.$$

Solution. By pecifying $\overline{AR}$, $\overline{SR} = 1$, the optimal solutions obtained are $d_{1,2}^{1,l} = d_{2,1}^{1,l} = 1$, and all others are $d_{j,p}^{1,l} = 0$, $u_1^{1,l} = v_2^{1,l} = v_3^{1,l} = v_4^{1,l} = v_5^{1,l} = 1$, $\lambda_1^{1,l} = \lambda_2^{1,l} = 0$, $\lambda_3^{1,l} = \lambda_4^{1,l} = 1$, $h^{1,l} = 2$. The objective value $CR^{1,l}$ is $\frac{1}{4}(2 + 1 - \frac{2-1}{12}) = 0.73$. The rule is exactly the same as $R^{1,1}$ in Table 2.3 and Table 2.4. □

Example 3.2. Similarly, the model of inducing a rule $R^{3,l}$ for class 3 is formulated below:

$$\text{Max } CR^{3,l}$$

subject to: equations (3.6)–(3.18) with changing superscripts $1, l$ to $3, l$, and

$$5(u_4^{3,l} - 1) + 4 - h^{3,l} \leq d_{1,1}^{3,l} + d_{2,1}^{3,l} + d_{3,4}^{3,l} + d_{42}^{3,l} \leq 4 - h^{3,l} + 5(1 - u_4^{3,l}),$$

$$5(u_5^{3,l} - 1) + 4 - h^{3,l} \leq d_{1,1}^{3,l} + d_{2,1}^{3,l} + d_{3,3}^{3,l} + d_{4,1}^{3,l} \leq 4 - h^{3,l} + 5(1 - u_5^{3,l}),$$

$$d_{1,2}^{3,l} + d_{2,1}^{3,l} + d_{3,2}^{3,l} + d_{4,3}^{3,l} \leq 4 - h^{3,l} - 1 + 5(1 - v_1^{3,l}),$$

$$d_{1,3}^{3,l} + d_{2,2}^{3,l} + d_{3,1}^{3,l} + d_{4,1}^{3,l} \leq 4 - h^{3,l} - 1 + 5(1 - v_2^{3,l}),$$

$$d_{1,1}^{3,l} + d_{2,2}^{3,l} + d_{3,2}^{3,l} + d_{4,3}^{3,l} \leq 4 - h^{3,l} - 1 + 5(1 - v_3^{3,l}),$$

$$AR^{3,l} = \frac{1}{5-2}(v_1^{3,l} + v_2^{3,l} + v_3^{3,l}) \geq \overline{AR},$$

$$SR^{3,l} = \frac{1}{2}(u_4^{3,l} + u_5^{3,l}) \geq \overline{SR},$$

$$v_1^{3,l}, v_2^{3,l}, v_3^{3,l}, u_4^{3,l}, u_5^{3,l}, d_{1,1}^{3,l}, d_{1,2}^{3,l}, d_{1,3}^{3,l}, d_{2,1}^{3,l}, d_{2,2}^{3,l}, d_{3,1}^{3,l}, d_{3,2}^{3,l}, d_{3,3}^{3,l}, d_{3,4}^{3,l},$$
$$d_{4,1}^{3,l}, d_{4,2}^{3,l}, d_{4,3}^{3,l}, \lambda_1^{3,l}, \lambda_2^{3,l}, \lambda_3^{3,l}, \lambda_4^{3,l} \in \{0,1\}.$$

Solution. By specifying $\overline{AR} = \overline{SR} = 1$, the optimal solutions obtained are $d_{3,3}^{3,l} = d_{3,4}^{3,l} = 1$, and all others are $d_{j,p}^{3,l} = 0$, $u_4^{3,l} = u_5^{3,l} = v_1^{3,l} = v_2^{3,l} = v_3^{3,l} = 1$, $\lambda_3^{3,l} = 0$, $\lambda_1^{3,l} = \lambda_2^{3,l} = \lambda_4^{3,l} = 1$, $h^{3,l} = 3$. The objective value $CR^{3,l}$ is $\frac{1}{4}(3 + 1 - \frac{2-1}{12}) = 0.98$. The rule is exactly the same as $R^{3,3}$ in Table 2.3 and Table 2.4. □

Model 3.2. (Specifying the aspiration levels for AR, SR, and CR)

$$\text{Max } AR^{k,l} + SR^{k,l} + CR^{k,l}$$

subject to: constraints (C1)–(C7), and

$$AR^{k,l} \geq \overline{AR},$$

$$SR^{k,l} \geq \overline{SR},$$

$$CR^{k,l} \geq \overline{CR}.$$

□

Model 3.3. (Specifying the weights on AR, SR and CR)

$$\text{Max } w_a^{k,l} AR^{k,l} + w_s^{k,l} SR^{k,l} + w_c^{k,l} CR^{k,l}$$

subject to: constraints (C1)–(C7), where $w_a^{k,l}$, $w_s^{k,l}$ and $w_c^{k,l}$ are the weighting value of $AR^{k,l}$, $SR^{k,l}$ and $CR^{k,l}$, respectively. □

In addition to inducing the best rule, we may also generate conveniently the second best, the third best, etc. rules.

Procedure 3.1. The solution procedure for Model 3.1 is:

Step 1. Specify the AR and SR

Step 2. Obtain the solution of Model 3.1

Step 3. If no feasible solution exists

Step 3.1. Decrease AR or SR, and go to Step 3

Else

Step 3.2. A rule is obtained

Step 4. If more rules are wanted,

Step 4.1. Add the solution obtained from Step 3.1 as a new constraint for Model 3.1, then go to Step 2.

In the first iteration of Step 3.2 in Procedure 3.1, we get the global optimal rule for a specified class; and in the second iteration, we get the second optimal rule, and so on. Example 3.3 illustrates the solution procedure using Procedure 3.1.

Example 3.3. Take Example 3.2 for instance. Since $u_4^{3,l} = u_5^{3,l} = 1$ is the solutions of Example 3.2, if one more rule for class 3 is needed, we can add the following new constraint

$$u_4^{3,l} + u_5^{3,l} < 2$$

to the model of Example 3.2. This constraint prevents $u_4^{3,l} = u_5^{3,l} = 1$, simultaneously. There is no feasible solution after adding the above constraint. It means that no more rule with $AR^{3,l} = SR^{3,l} = 1$ can be induced. Then, we can decrease the acceptable level of $\overline{AR}$ or $\overline{SR}$ to get second best rules. Here, we decrease $\overline{SR}$ to 0.5, and the solutions obtained are $d_{3,2}^{3,l} = 1$; all others are: $d_{j,p}^{3,l} = 0$, $u_4^{3,l} = 0$, $u_5^{3,l} = v_1^{3,l} = v_2^{3,l} = v_3^{3,l} = 1$, $\lambda_3^{3,l} = 0$, $\lambda_1^{3,l} = \lambda_2^{3,l} = \lambda_4^{3,l} = 1$, $h^{3,l} = 3$, $AR^{3,l} = CR^{3,l} = 1$, $SR^{3,l} = 0.5$. The rule is exactly the same as

$R^{3,1}$ in Table 2.3, which is the second best rule for class 3. By adding the next constraint,

$$u_5^{3,l} < 1$$

to the model, the third best rule, exactly the same as $R^{3,2}$ in Table 2.3, is then obtained. $\square$

3.4 Analysis of Models

For a data set having n objects and characterized by m attributes with q sub-attributes, the analysis of constraints and binary variables for inducing a specific rule $R^{k,l}$ is described below.

- For each object, it needs either a constraint of (C1) or a constraint of (C2). So the instance of constraints of (C1) and (C2) is n.
- For each sub-attribute, it needs a constraint of (C3). So the instance of constraints of (C3) is q.
- For each attribute, it needs a constraint of (C4). So the instance of constraints of (C4) is m.
- There is only one instance of constraint (C5).
- The instance of constraints of (C6) depends on each q_j. The worst case is only one $q_j \neq 1$ and the instance of constraints of (C6) is $q - m - 2$.
- The number of binary variables $u_i^{k,l}$ and $v_r^{k,l}$ is n, since each object needs a $u_i^{k,l}$ or $v_r^{k,l}$.
- To represent a rule, it needs a binary variable $d_{j,p}^{k,l}$ for each sub-attribute. So the number of $d_{j,p}^{k,l}$ is q.

- The number of binary variables $\lambda_j^{k,l}$ is m.

To sum up, the maximum number of constraints of (C1)–(C6) is $n + 2q - 1$ and the total number of binary variables is $n + q + m$.

Chapter 4 Experiments

This chapter demonstrates the solution process of the proposed method by two practical data sets, the HSV (Highly Selective Vagotomy) patients and the European Barn Swallow, and compares the induction results with RST (or VPRS) and ID3.

4.1 The HSV Patients Data Set

The HSV patients data set is a clinical data set of F. Raszeja Mem. Hospital in Poland. HSV, also called proximal gastric vagotomy, is an effective method of treatment of duodenal ulcer, which consists of vagal denervating of the stomach area secreting hydrochloric acid (Dunn et al., 1980). This data set is composed of 122 patients with duodenal ulcer treated by HSV, as described by 11 pre-operating attributes. For more details about the data set, please see Appendix A. The patients are classified into four classes, according to a long-term result of HSV, and all evaluated by a surgeon in the modified Visick grading, following the definition of Goligher et al. (1978). Exact values of the considered quantitative attributes are translated into ordered qualitative terms, i.e., "high," "medium," "low," etc. This translation is due to some empirical norms defining intervals of attribute values corresponding to qualitative terms. The terms are then coded by numbers 1, 2, 3, etc., which create the domain of coded attributes. The norms adopted in the study are shown in Table 4.1 (Slowinski, 1992).

Table 4.1: Norms for attributes of the HSV data set

No.	Attribute [units]	Domain (code) 1	2	3	4	5	Remarks
a_1	Gender	male	female				
a_2	Age [years]	≤ 35	> 35				
a_3	Duration of disease [years]	≤ 0.5	$(0.5, 3]$	> 3			
a_4	Complication of ulcer	none	acute haemorrhage	multiple haemorrhage	perforation in the past	pyloric stenosis	
a_5	HCL concentration [mmol HCL/100ml]	≤ 2	$(2, 4]$	> 4			basic secretion
a_6	Volume of gastric juice per 1h [ml]	≤ 70	$(70, 150]$	> 150			
a_7	Volume of residual gastric juice [ml]	≤ 50	$(50, 100]$	> 100			
a_8	Basic acid output (BAO) [mmol HCL/h]	≤ 2	$(2, 3]$	> 3			
a_9	HCL concentration [mmol HCL/100ml]	≤ 10	$(10, 15]$	> 15			secretion stimulated
a_{10}	Volume of gastric juice per 1h [ml]	≤ 100	$(100, 250]$	> 250			by histamine
a_{11}	Maximal acid output [mmol HCL/h]	≤ 15	$(15, 25]$	$(25, 40]$	> 40		

This table is extracted from Slowinski (1992).

Table 4.2: The best rule for each class of HSV found by ROSE2

No.	Rule
1	if $(a_2 = 2) \wedge (a_4 = 1) \wedge (a_6 = 2)$, then $c = 1$
2	if $(a_1 = 1) \wedge (a_2 = 1) \wedge (a_4 = 1) \wedge (a_5 = 3) \wedge (a_9 = 2) \wedge (a_{11} = 3)$, then $c = 2$
3	if $(a_3 = 1) \wedge (a_4 = 1)$, then $c = 3$
4	if $(a_4 = 1) \wedge (a_6 = 3) \wedge (a_{10} = 3) \wedge (a_{11} = 4)$, then $c = 4$

4.1.1 Rules Induced by RST and ID3

Here, we use a software tool named ROSE2 (Rough Sets Data Explorer)(Predki et al., 1998; Predki and Wilk, 1999) to induce rules by RST. The computations of ROSE2 are based on rough-set fundamentals. Table 4.2 lists the best rule for each class found by ROSE2.

Figure 4.1 is the partial HSV classification tree by ID3, which only lists the best path for each class. For example, from the branch $a_4 = 1$, $a_6 = 3$, $a_{11} = 4$ and $a_{10} = 3$, the leaf $c = 4$ is reached. It means that

"*if* $(a_4 = 1)$ *and* $(a_6 = 3)$ *and* $(a_{11} = 4)$ *and* $(a_{10} = 3)$,

then the objects belong to class 4".

4.1.2 Rules Induced by the Proposed Method

To simplify the presentation, we utilize Model 3.1 to induce rules where the accuracy rate is fixed at 100%. The used model is as follows:

$$\text{Max } CR^{k,l}$$

subject to: constraints (C1)–(C7), and

$$AR^{k,l} = 1,$$

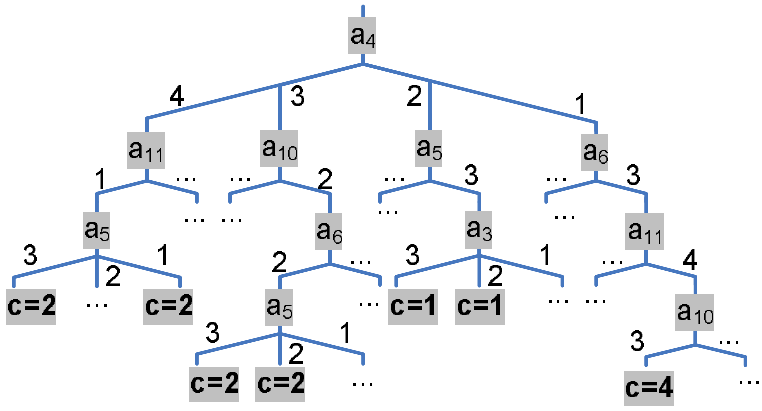

Figure 4.1: A partial ID3 decision tree for HSV data set.

$$SR^{k,l} \geq \alpha,$$

where $CR^{k,l}$ is to be maximized with restrictions that $AR^{k,l} = 1$ and $SR^{k,l} \geq \alpha$. α is a parameter value. The best rule found and the second best rules for each class are listed in Table 4.3. $R^{1,1}$ means the best rule for class 1, and $R^{1,2}$ is the second best rule for class 1. $R^{1,1}$ is found by specifying $\alpha = \frac{19}{79}$ (there is no feasible solution for $\alpha > \frac{19}{79}$), which means

"*if* $(a_2 = 2)$ *and* $(a_4 = 1$ *or* 2 *or* $3)$ *and* $(a_9 = 3)$,

then the objects belong to class 1".

The rule $R^{1,1}$ is supported by 19 objects. They are x_1, x_4, x_9, x_{11}, x_{15}, x_{19}, x_{25}, x_{27}, x_{46}, x_{57}, x_{61}, x_{71}, x_{83}, x_{88}, x_{94}, x_{104}, x_{106}, x_{111}, and x_{117}. Its support rate is $SR^{1,1} = \frac{19}{79} = 0.24$. There are eight attributes ignored and five active sub-attributes in $R^{1,1}$, so its compactness rate is $CR^{1,1} = \frac{1}{11}(8 + 1 - \frac{5-1}{34}) = 0.81$. The second best rule $R^{1,2}$ is conveniently induced by adjusting $\alpha = \frac{17}{79}$ and adding the following new constraint to Model 3.1.

$$u_1^{1,2} + u_4^{1,2} + u_9^{1,2} + u_{11}^{1,2} + u_{15}^{1,2} + u_{19}^{1,2} + u_{25}^{1,2} + u_{27}^{1,2} + u_{46}^{1,2} + u_{57}^{1,2} + u_{61}^{1,2}$$
$$+ u_{71}^{1,2} + u_{83}^{1,2} + u_{88}^{1,2} + u_{94}^{1,2} + u_{104}^{1,2} + u_{106}^{1,2} + u_{111}^{1,2} + u_{117}^{1,2} < 19.$$

The solution is $R^{1,2}$, which means

"*if* $(a_4 = 2)$ *and* $(a_9 = 2$ *or* $3)$, *then the objects belong to class* 1"

with $CR^{1,2} = 0.90$ and $SR^{1,2} = 0.16$. In the same processes, the best rules for remaining three classes are obtained. The SR of the best rules for classes 2, 3, 4 are 0.22, 0.38,0.23, respectively, and CR are 0.72, 0.81, 0.63, respectively

4.1.3 Comparison of Results

Here, we compare the proposed method with ROSE2 and ID3. Table 4.3 lists the best rules found by the three methods. All the rules listed here are 100% accuracy. For class 1, $R^{1,1}$ with $CR^{1,1} = 0.81$ and $SR^{1,1} = 0.24$ and $R^{1,2}$ with $CR^{1,2} = 0.90$ and $SR^{1,2} = 0.16$ are the best and second best rules found by the proposed method, respectively. $R^{1,3}$ with $CR^{1,3} = 0.81$ and $SR^{1,3} = 0.16$ is the best rule that can be found by ROSE2. It is worse than $R^{1,1}$ and $R^{1,2}$. $R^{1,4}$ with $CR^{1,4} = 0.81$ and $SR^{1,4} = 0.15$, the best rule that can be found by ID3, is even worse than $R^{1,3}$. For class 2, $R^{2,1}$ with $CR^{2,1} = 0.72$ and $SR^{2,1} = 0.22$, the best rule found by both the proposed method and ID3, is better than $R^{2,2}$ with $CR^{2,2} = 0.53$ and $SR^{2,2} = 0.22$, the best rule that can be found by ROSE2. Consider class 3, $R^{3,1}$ with $CR^{3,1} = 0.81$ and $SR^{3,1} = 0.38$ is the best rule found by the proposed method. Although $R^{3,2}$ with $CR^{3,2} = 0.91$ and $SR^{3,2} = 0.25$ is the second best rule found by the proposed method, it is the best rule that can be found by ROSE2. $R^{3,3}$ with $CR^{3,3} = 0.81$ and $SR^{3,3} = 0.25$, the best rule that can be found by ID3, is worse than $R^{3,2}$. Similarly, $R^{4,1}$, the best rule found by proposed method, is better than $R^{4,3}$, the best rule that can be found by ROSE2 and ID3.

Table 4.3: Comparison of the proposed method, ROSE2 and ID3 for the HSV data set.

Class	Used methods	Induced rule	Meaning of rule	AR	SR	CR	Supporting objects
1	Proposed	$R^{1,1}$	if $(a_2 = 2) \wedge (a_4 = 1 \vee 2 \vee 3) \wedge (a_9 = 3)$, then $c = 1$	1	0.24	0.81	x_1, x_4, x_9, x_{11}, x_{15}, x_{19}, x_{25}, x_{27}, x_{46}, x_{57}, x_{61}, x_{71}, x_{83}, x_{88}, x_{94}, x_{104}, x_{106}, x_{111}, x_{117}
	Proposed	$R^{1,2}$	if $(a_4 = 2) \wedge (a_9 = 2 \vee 3)$, then $c = 1$	1	0.16	0.90	x_2, x_6, x_7, x_8, x_{16}, x_{46}, x_{47}, x_{52}, x_{56}, x_{57}, x_{88}, x_{116}, x_{117}
	ROSE2	$R^{1,3}$	if $(a_2 = 2) \wedge (a_4 = 1) \wedge (a_6 = 2)$, then $c = 1$	1	0.16	0.81	x_1, x_9, x_{19}, x_{25}, x_{31}, x_{48}, x_{67}, x_{83}, x_{93}, x_{97}, x_{106}, x_{114}, x_{115}
	ID3	$R^{1,4}$	if $(a_3 = 2 \vee 3) \wedge (a_4 = 2) \wedge (a_5 = 3)$, then $c = 1$	1	0.15	0.81	x_2, x_6, x_7, x_8, x_{16}, x_{46}, x_{47}, x_{52}, x_{56}, x_{88}, x_{116}, x_{117}
2	Proposed / ID3	$R^{2,1}$	if $(a_4 = 3) \wedge (a_5 = 2 \vee 3) \wedge (a_6 = 2) \wedge (a_{10} = 2)$, then $c = 2$	1	0.22	0.72	x_{12}, x_{45}, x_{62}, x_{110}
	ROSE2	$R^{2,2}$	if $(a_1 = 1) \wedge (a_2 = 1) \wedge (a_4 = 1) \wedge (a_5 = 3) \wedge$ $(a_9 = 2) \wedge (a_{11} = 3)$, then $c = 2$	1	0.22	0.53	x_{87}, x_{90}, x_{103}, x_{120}
3	Proposed	$R^{3,1}$	if $(a_3 = 2) \wedge (a_9 = 1) \wedge (a_{11} = 2 \vee 3)$, then $c = 3$	1	0.38	0.81	x_{43}, x_{79}, x_{109}
	Proposed / ROSE2	$R^{3,2}$	if $(a_3 = 1) \wedge (a_4 = 1)$, then $c = 3$	1	0.25	0.91	x_5, x_{56}
	ID3	$R^{3,3}$	if $(a_4 = 4) \wedge (a_5 = 1 \vee 3) \wedge (a_{11} = 1)$, then $c = 3$	1	0.25	0.81	x_{13}, x_{24}
4	Proposed	$R^{4,1}$	if $(a_1 = 2) \wedge (a_2 = 2) \wedge (a_3 = 3) \wedge (a_5 = 3) \wedge$ $(a_6 = 1)$, then $c = 4$	1	0.23	0.63	x_{18}, x_{75}, x_{107}
	Proposed	$R^{4,2}$	if $(a_6 = 3) \wedge (a_9 = 3) \wedge (a_{10} = 3)$, then $c = 4$	1	0.15	0.81	x_{39}, x_{75}
	ROSE2 / ID3	$R^{4,3}$	if $(a_4 = 1) \wedge (a_6 = 3) \wedge (a_{10} = 3) \wedge (a_{11} = 4)$, then $c = 4$	1	0.15	0.72	x_{39}, x_{75}

The rules are ranked by SR then CR.

Table 4.4: Norms for attributes of the European barn swallow data set

		Domain (code)			
No.	Attribute [units]	1	2	3	4
a_1	Head and bill length [mm]	$[28.9, 29.48)$	$[29.48, 30.05)$	$[30.05, 30.53)$	$[30.53, 31)$
a_2	Right streamer length [mm]	$[74, 87.8)$	$[87.8, 101.5)$	$[101.5, 115.3)$	$[115.3, 129)$
a_3	Mid tail length [mm]	$[41, 43.8)$	$[43.8, 46.5)$	$[46.5, 48.3)$	$[48.3, 50)$
a_4	Left streamer length [mm]	$[77, 89.8)$	$[89.8, 102.5)$	$[102.5, 115.8)$	$[115.8, 129)$
a_5	Right wing length [mm]	$[118, 121.3)$	$[121.3, 124.5)$	$[124.5, 128.8)$	$[128.8, 133)$
a_6	Left wing length [mm]	$[118, 121.3)$	$[121.3, 124.5)$	$[124.5, 128.8)$	$[128.8, 133)$
a_7	Mass [g]	$[16.6, 18.38$	$[18.38, 20.15)$	$[20.15, 22.63)$	$[22.63, 25.1)$
a_8	Wing area [mm^2]	$[3780, 4873)$	$[4873, 5966)$	$[5966, 6291)$	$[6291, 6616)$

4.2 The European Barn Swallow Data Set

The European barn swallow (*Hirundo rustica*) data set was obtained by trapping individual swallows in Stirlingshire, Scotland, between May and July 1997 (Beynon and Buchanan, 2003). This data set contains 69 swallows, which were described by eight attributes. For more details about the data set, please see Appendix B. The birds are classified by gender of each bird. The norms of attributes of the swallow data set are shown in Table 4.4.

4.2.1 Rules Found by VPRS and ID3

The rules found by VPRS (variable precision rough sets), a variant of RST, as shown in Beynon and Buchanan (2003), are listed in Table 4.5, where AR, SR and CR values are computed by this study.

The rules found by ID3, also appearing in Beynon and Buchanan (2003), are listed in Figure 4.2. Each terminal node of the decision tree classifies all its associated objects correctly into the identified decision class. At each node,

Table 4.5: Rules for swallow data set found by VPRS

	Rule	AR	SR	CR
1	if $(a_2 = 1 \vee 2) \wedge (a_5 = 1 \vee 2) \wedge (a_6 = 1 \vee 2)$, then $c = 1$	0.97	0.66	0.730
2	if $(a_1 = 3 \vee 4) \wedge (a_2 = 1 \vee 2)$, then $c = 1$	1	0.16	0.863
3	if $(a_2 = 3 \vee 4)$, then $c = 2$	1	0.72	0.996
4	if $(a_1 = 1 \vee 2) \wedge (a_5 = 3 \vee 4) \wedge (a_6 = 3 \vee 4)$, then $c = 2$	0.97	0.31	0.730

The rules are extracted from Beynon and Buchanan (2003) where AR, SR and CR are computed by this study.

Table 4.6: Some better rules for swallow data set found by ID3

	Rule	AR	SR	CR
1	if $(a_2 = 1 \vee 2) \wedge (a_5 = 3 \vee 4) \wedge (a_3 = 2 \vee 3 \vee 4)$, then $c = 1$	1	0.66	0.727
2	if $(a_2 = 2) \wedge (a_5 = 3 \vee 4) \wedge (a_1 = 2 \vee 3 \vee 4)$, then $c = 1$	1	0.06	0.730
3	if $(a_2 = 3 \vee 4)$, then $c = 2$	1	0.72	0.996
4	if $(a_2 = 1 \vee 2) \wedge (a_5 = 1 \vee 2) \wedge (a_3 = 1)$, then $c = 2$	1	0.03	0.734
5	if $(a_2 = 1 \vee 2) \wedge (a_5 = 3 \vee 4) \wedge (a_1 = 1)$, then $c = 2$	1	0.16	0.734

information is given for the objects associated with it. From the root (top) node the left-hand branch is defined by $a_2 \leq 101.5$ with $41[32, 9]$ representing 41 objects satisfying this criteria. Of them, 32 objects have class value '1' and nine objects have class value '2'. Also shown to the right of every node box is the majority proportion of the objects in a same decision class. Continuing the same example as before, $32/41 = 0.780$ is the highest proportion of the objects at that node to the same decision class. Since this is a complete tree, a terminal node is identified when the associated majority proportion value equals 100%. Figure 4.2 can be expressed in "If $\cdots$ Then $\cdots$" form as shown in Table 4.6.

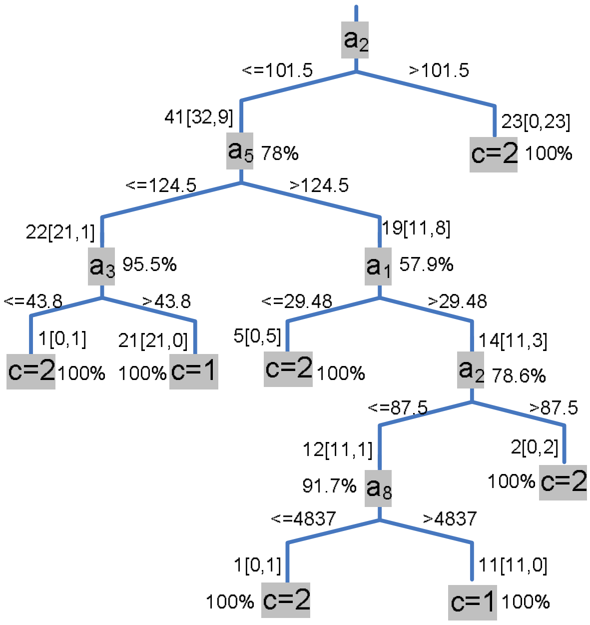

Figure 4.2: The ID3 decision tree for swallow data set. (Beynon and Buchanan, 2003)

4.2.2 Rules Induced by the Proposed Method

Here, we use Model 3.2 to induce rules. First, we specify $\overline{AR} \geq 0.97$, $\overline{SR} \geq 0.66$ and $\overline{CR} \geq 0.730$, then the optimal solution is the first rule in Table 4.7. Similarly, we specify other three sets of α values to get other three rules as listed in Table 4.7.

Table 4.7: Rules for swallow data set found by the proposed method

	Rule	$\overline{AR}$	$\overline{SR}$	$\overline{CR}$	AR	SR	CR
1	if $(a_2 = 1 \vee 2) \wedge (a_5 = 1 \vee 2)$, then $c = 1$	0.97	0.66	0.730	0.97	0.66	0.863
2	if $(a_1 = 4) \wedge (a_2 = 1 \vee 2)$, then $c = 1$	1	0.16	0.863	1	0.16	0.867
3	if $(a_2 = 3 \vee 4)$, then $c = 2$	1	0.72	0.996	1	0.72	0.996
4	if $(a_1 = 1 \vee 2) \wedge (a_6 = 3 \vee 4)$, then $c = 2$	0.97	0.31	0.730	0.97	0.31	0.863

4.2.3 Comparison of Results

Compare Table 4.7 with Table 4.5 and Table 4.6 to know that the proposed method can induce rules with better or equivalent values of AR, SR and CR. For example, the aspiration levels $\overline{AR}$, $\overline{SR}$ and $\overline{CR}$ of the four rules in Table 4.7 are set as equal to the corresponding AR, SR and CR in Table 4.5. The results show that the first, second and fourth rules in Table 4.7 are better than the corresponding rules in Table 4.5, and the third rule in both table are the same. In fact, the proposed method can induce optimal solutions while ROSE2 and ID3 may find only feasible solutions.

Chapter 5 Implementation

MCOCR (Multiple Criteria Optimal Classification Rules) is a software tool implementing the models proposed by this study. It is currently a prototype. MCOCR is developed by **DELPHI 7.0** and runs on the **Microsoft WindowsXP** operating system. The computing kernel behind MCOCR is **LINGO 9.0**. MCOCR gathers input data and relative parameters from users in order to generate a corresponding **LINGO** program, then calls **LINGO** to solve such a program. Finally, MCOCR interprets the results returned by **LINGO** into rules form.

Before starting MCOCR, an input data file must be prepared. The input data file contains the data set to be induced rules. In addition, the input data file also contains some meta information about the data set itself. So it must follow the input data file format, otherwise MCOCR cannot work properly. For more detail about input data file format, please see Appendix C.

When starting MCOCR, you will see the window shown in Figure5.1. There are five tags in the window:

- **Model**: for choosing which model (i.e., Model 3.1, 3.2 or 3.3) used
- **Parameter**: for inputting parameters that MCOCR needs
- **Origin Data**: for showing contents of an input data file
- **Program**: for showing the LINGO model generated by MCOCR
- **Result**: for showing the rules induced by MCOCR

So, the fist thing is to decide what model used. Here, Model 3.1 is used as an example. After chosen a model, the user must input some related parameters. There are some steps to induce rules.

- Select an input data file (Figure 5.1). After chosen an input data file, the "Origin Data" tag appears (Figure 5.2). Click the tag, contents of the input file will be displayed on the window (Figure 5.3).
- Select an objective, for instance, "max CR" (Figure 5.4).
- Specify lower bounds, for instance, "AR" and "number of supporting objects" (Figure 5.4).
- Specify the class value to classify (Figure 5.4).
- Generate **LINGO** program (Figure 5.4). While **LINGO** program generated, the "Program" tag appears (Figure 5.5). Click the tag, contents of the generated **LINGO** program will be displayed on the window (Figure 5.6).
- Induce rules(Figure 5.7). The "Result" tag will appear while a rule generated and generated rules will be displayed on the window (Figure 5.8).
- Induce next rules if needed (Figure 5.9). The rules will display on the window after previous rules (Figure 5.10).

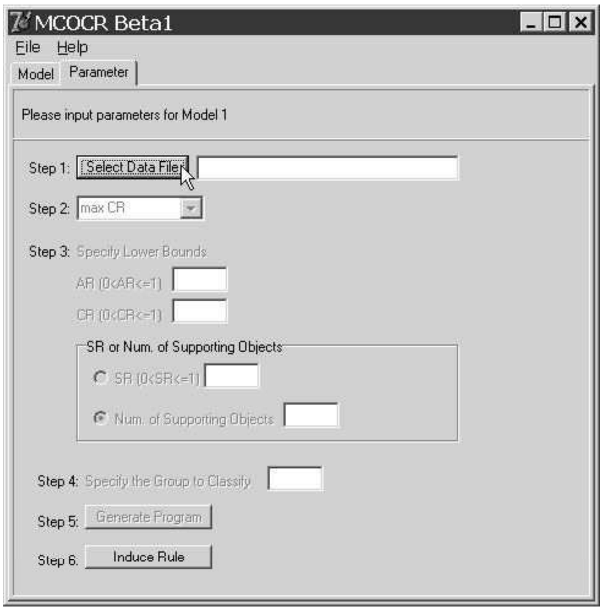

Figure 5.1: Step1: Click "Select Data File" button to select an input data file

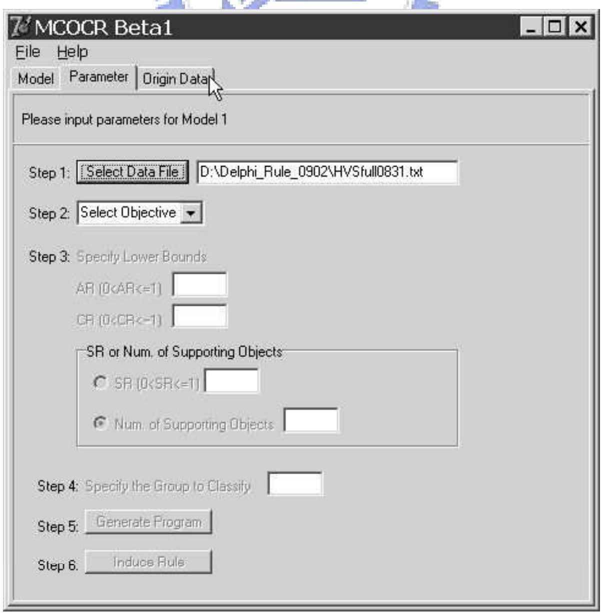

Figure 5.2: After chosen an input data file, the "Origin Data" tag appears.

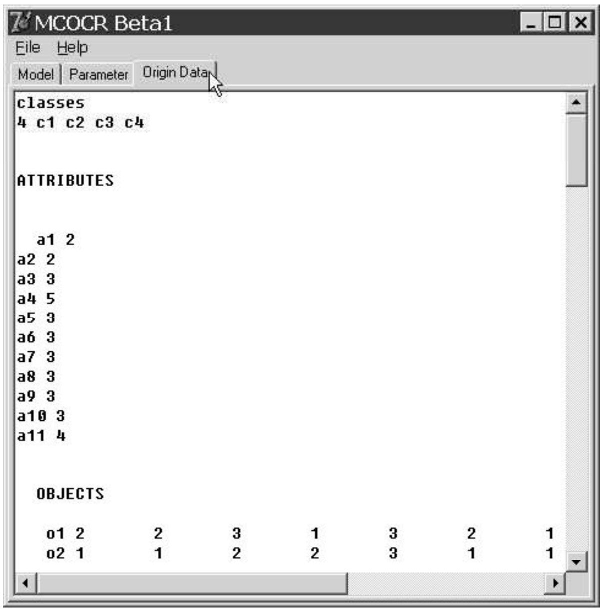

Figure 5.3: Click "Origin Data" tag, the contents of the input file will be displayed on the window.

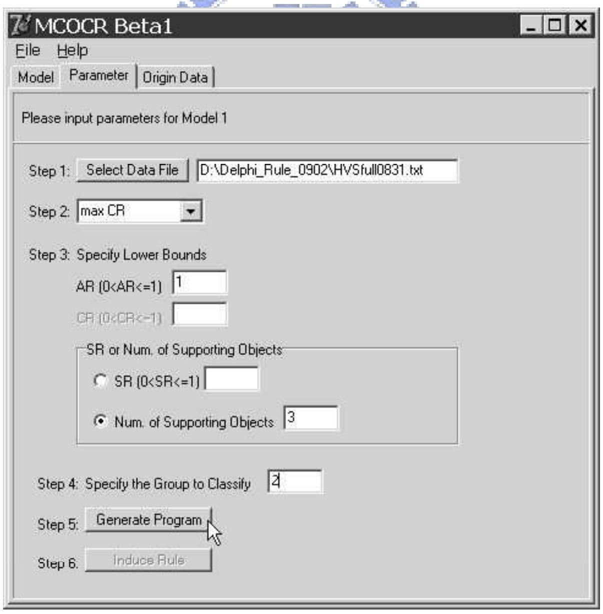

Figure 5.4: Select an objective. The "max CR" is chosen, here. Specify Lower Bounds. The AR is specified as 1 and the number of supporting objects is specified as 3, here. Specify the class to classify. Here is 2. Click "Generate Program" button to generate **LINGO** program.

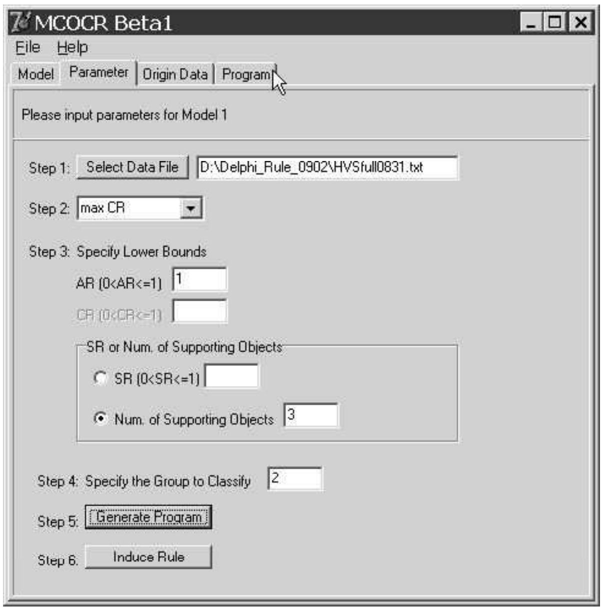

Figure 5.5: After "Generate Program" button clicked, the "Program" tag appears.

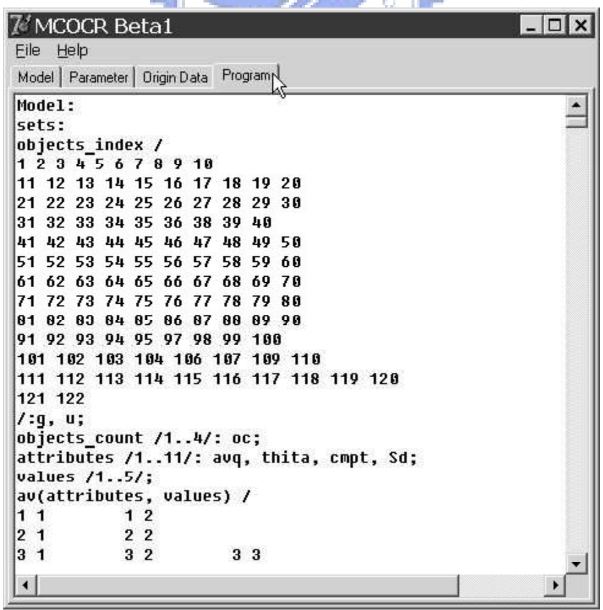

Figure 5.6: Click "Program" tag, the contents of the generated Lingo program will be displayed on the window.

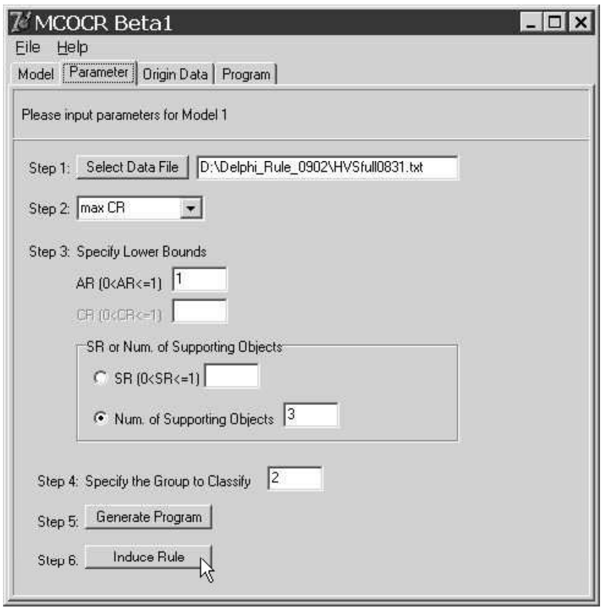

Figure 5.7: Step 6: Click "Induce Rule" button to start inducing rules.

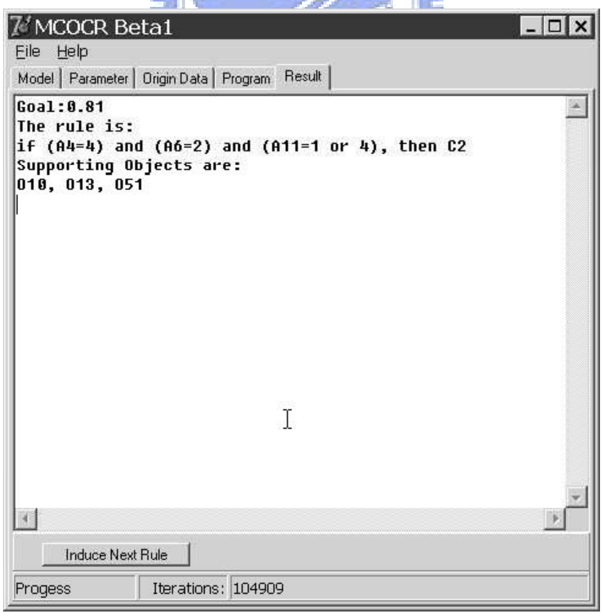

Figure 5.8: The "Result" tag will appear while a rule generated and generated rules will be displayed on the window.

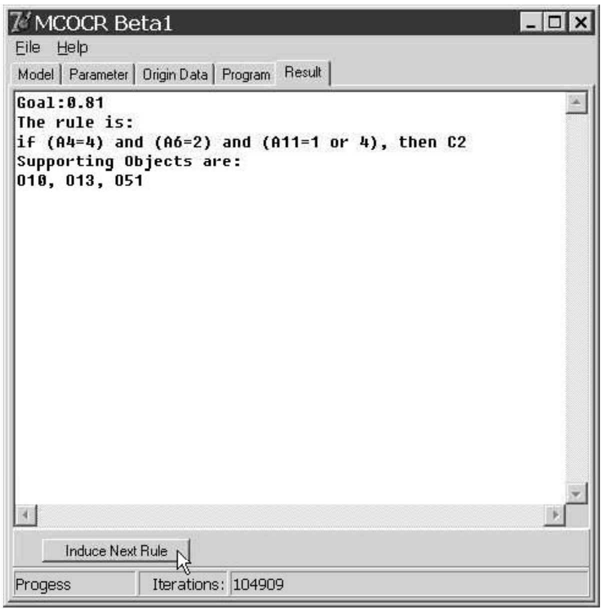

Figure 5.9: Click "Induce Next Rule" button to induce another rule.

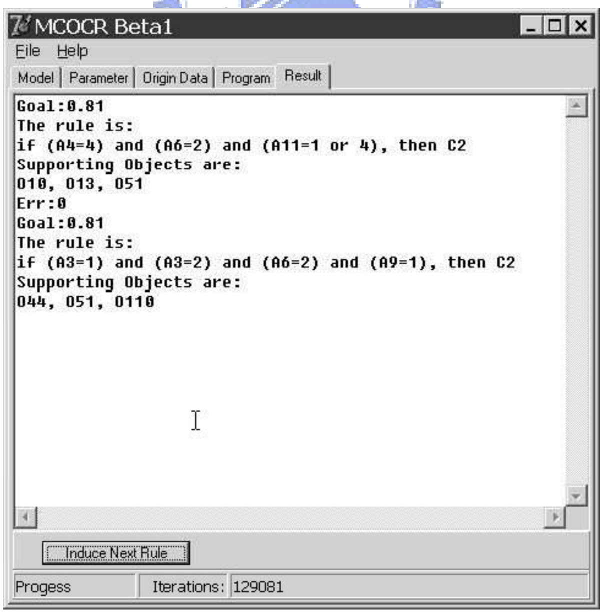

Figure 5.10: The result of "Induce Next Rule".

Chapter 6 Discussions and Remarks

6.1 Discussions

This study develops a multiple criteria mixed 0-1 linear programming model to induce rules. Some advantages of the proposed method are listed below:

(i) Solution quality: The rules obtained by the proposed method are globally optimal solutions, but the rules obtained by RST or ID3 may just be feasible solutions.

(ii) Multiple criteria: Three criteria accuracy rate, support rate, and compact rate are considered to be maximized simultaneously.

(iii) Constraints: The proposed method is conveniently to add other constraints to fit requirements, but it is difficult to do in RST and ID3 methods.

Although the proposed method is applied in biology and medicine informatics, it can apply in a variety of research and application areas.

6.2 Remarks

Although the adventages mentioned above, there are some limitations of proposed method, which are the future works of this study.

The number of binary variables $u_i^{k,l}$ and $v_r^{k,l}$ is direct propotion to the number of obejcts n in such data set. While the number of objects become large, the computation time will increase, seriously. The numbers of binary variables $\lambda_j^{k,l}$ and $d_{j,p}^{k,l}$ are also direct propotion to the numbers of attributes m

and sub-attributes q, respectively, but this is a relatively minor problem since the numbers of attributes and sub-attributes are not very large in most cases. So, how to discrease the number of $u_i^{k,l}$ and $v_r^{k,l}$ is a major issue for future works.

References

Aja-Fernandez, S., Luis-Garcia, R., Martin-Fernandez, M. A., and Alberola-Lopez, C. (2004), "A computational TW3 classifier for skeletal maturity assessment — a computing with words approach", *Journal of Biomedical Informatics*, 37, 99–107.

Beynon, M. J. and Buchanan, K. L. (2003), "An illustration of variable precision rough set theory: The gender classification of the European Barn Swallow (*Hirundo rustica*)", *Bulletin of Mathematical Biology*, 65, 835–858.

Dunn, D. C., Thomas, W. E. G., and Hunter, J. O. (1980), "An evaluation of highly selective vagotomy in the treatment of chronic duodenal ulcer", *Surgery Gynecology Obstetrics*, 150, 845–849.

Geurts, P., Fillt, M., Seny, D., Meuwis, M.-A., Malaise, M., Merville, M.-P., and Wehenkel, L. (2005), "Proteomic mass spectra classification using decision tree based ensemble methods", *Bioinformatics*, 21, 3138–3145.

Goligher, J. C., Hill, G. L., Kenny, T. E., and Nutter, E. (1978), "Proximal gastric vagotomy without drainage for duodenal ulcer: results after 5-8 years", *British Journal Surgery*, 65, 145–151.

Hvidsten, T. T., Lægreid, A., and Komorowski, J. (2003), "Learning rule-based models of biological process from gene expression time profiles using gene ontology", *Bioinformatics*, 19, 1116–1123.

Kay, A. (1967), "Memorial lecture — an evaluation of gastric acid secretion tests", *Gastroenterology*, 53, 834–852.

Li, H.-L. and Fu, C.-J. (2005), "A linear programming approach for identifying a consensus sequence on DNA sequences", *Bioinformatics*, 21, 1838–1845.

Li, R. and Wang, Z.-O. (2004), "Mining classification rules using rough sets and neural networks", *European Journal of Operational Research*, 157, 439–448.

Pawlak, Z. (1982), "Rough sets", *International Journal of Information and Computer Sciences*, 11, 341–356.

Predki, B., Slowinski, R., Stefanowski, J., Susmaga, R., and Wilk, S. (1998), "ROSE — software implementation of the rough set theory", in L. Polkowski and A. Skowron (eds.), *Rough Sets and Current Trends in Computing*, volume 1424 of *Lecture Notes in Computer Science*, 605–608, Berlin: Springer.

Predki, B. and Wilk, Sz. (1999), "Rough set based data exploration using ROSE system", in Z. W. Ras and A. Skowron (eds.), *Foundations of Intelligent Systems*, volume 1609 of *Lecture Notes in Computer Science*, 172–180, Berlin: Springer.

Quinlan, J. R. (1986), "Induction of decision trees", *Machine Learning*, 1, 81–106.

——— (1993), *C4.5: Programs for Machine Learning*, Los Altos, CA: Morgan Kaufmann.

Shen, Lixiang and Loh, Han Tong (2004), "Applying rough sets to market timing decisions", *Decision Support Systems*, 37, 583–598.

Slowinski, K. (1992), "Rough classification of HSV patients", in R. Slowinski (ed.), *Intelligent Decision Support — Handbook of applications and advances of the rough sets theory*, 77–94, Dordrecht, Nethrelands: Kluwer.

Sun, M. and Xiong, M. (2003), "A mathematical programming approach for gene selection and tissue classification", *Bioinformatics*, 19, 1243–1251.

Tay, Francis E. H. and Shen, Lixiang (2002), "Economic and financial prediction using rough sets model", *European Journal of Operational Research*, 141, 641–659.

Tsumoto, R. (1999), "Discovery of rules for medical expert systems — rough set approach", in *Proceedings of Third International Conference on Computational Intelligence and Multimedia Applications*, 212–216, ICCIMA'99.

Zhang, H., Yu, C.-Y., Singer, B., and Xiong, M. (2001), "Recursive partitioning for tumor classification with gene expression microarray data", in *Proceedings of the National Academy of Sciences of the USA*, volume 98, 6730–6735.

Ziarko, W. (1993), "Variable precision rough set model", *Journal of Computer and System Sciences*, 46, 39–59.

Appendices

A The HSV Patients Data Set

The data set as shown in Table A.1, is composed of 122 patients with duodenal ulcer treated by HVS, described by 11 pre-operating attributes. Attribute 1 – 4 concern anamnesis, and the remaining attributes are related to pre-operation gastric secretion examined with the histaminic test of Kay (1967). The patients are classified according to a long term result of HVS, evaluated by a surgeon in the modified Visick grading. The grading was derived from the following definition Goligher et al. (1978):

- **Excellent**: absolutely no symptoms, perfect result. The class index, 1, is given.
- **Very good**: patient considers result perfect, but interrogation elicits mild occasional symptoms easily controlled by a minor adjustment of diet. The class index, 2, is given.
- **Satisfactory**: mild or moderate symptoms easily controlled by care, which cause some discomfort, but patient and surgeon are satisfied with result which dose not interfere seriously with life or work. The class index, 3, is given.
- **Unsatisfactory**: moderate or sever symptoms of complications which interfere with work or normal life; patient or surgeon dissatisfied with result; includes all cases with recurrent ulcer and those submitted to further operation, even though the latter may have been followed by considerable symptomatic improvement. The class index, 4, is given.

Table A.1: The HSV patients data set with original values of attributes

No.	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11	class
1	1	46	12	0	5.6	79	50	4.4	19	119	22.6	1
2	0	27	3	1	12.5	58	15	7.3	26	120	31.2	1
3	0	25	6	0	11.5	77	15	8.9	16.1	93	15	1
4	0	48	3	0	15.6	29	2	4.5	28.7	186	53.4	1
5	1	26	0.5	0	7.6	80	45	6.1	17.1	101	17.2	3
6	0	32	5	1	11.9	56	100	6.7	13.6	150	20.4	1
7	0	26	2	1	6.1	19	8	1.2	14.8	58	8.6	1
8	1	28	2	1	6	36	40	2.2	20.4	65	13.33	1
9	0	55	30	0	16.8	118	12	19.8	40.4	172	69.6	1
10	0	21	5	3	20.9	111	32	23.2	34.5	270	93.1	2
11	0	37	2	0	12.6	152	30	19.2	38.7	202	78.2	1
12	0	48	5	2	2.3	73	6	1.7	5.5	199	10.9	2
13	0	43	20	3	8.1	97	32	7.8	11	120	13.2	3
14	0	30	2	0	10	15	15	1.5	18.8	121	22.7	1
15	0	49	14	2	11.7	118	38	13.8	23.2	266	52.5	1
16	0	27	3	1	9.5	154	25	14.6	13.5	141	19.1	1
17	0	28	10	0	20.9	178	26	36.1	23.3	214	49.8	1
18	1	40	4	0	8.1	62	17	5	5.6	41	2.3	4
19	0	60	20	0	13.4	107	27	14.3	19	335	63.5	1
20	0	22	4	0	3.5	176	40	6.1	5.6	190	10.6	2
21	0	21	4	0	1	155	66	1.6	2.6	160	4.2	1
22	0	21	6	4	4	360	210	14.4	3.4	211	7.1	1
23	0	28	0	1	6	152	15	9.2	9.8	227	22.3	1
24	0	31	2	3	1.8	60	10	1.1	12.3	117	14.4	3
25	0	37	3	0	8.5	94	20	8	17.3	188	32.6	1
26	0	22	2	0	8.3	111	28	9.2	20.8	192	39.8	1
27	0	43	5	0	1.9	401	53	7.5	16.3	94	15.2	1
28	1	59	1	0	4.8	30	12	1.4	9.3	27	5.2	1
29	0	32	3	0	2.8	164	35	4.5	10.3	178	18.3	1
30	0	34	8	0	6.3	82	13	5.2	7.4	130	9.6	1
31	0	51	1	0	8.6	87	25	7.5	13.7	230	31.4	1
32	0	41	20	0	2.6	29	15	0.8	6.1	108	6.6	1
33	1	50	5	1	2.5	44	120	1.1	4.1	49	2.1	1
34	0	24	2	0	14.1	160	22	22.5	21.2	209	44.4	1
35	0	32	3	0	9	122	45	10.9	15.7	223	35	1
36	0	30	8	0	8.5	121	26	10.3	5.7	261	11.4	1
37	0	63	2	0	5.8	60	34	3.5	8.7	133	11.5	1
38	0	30	2	1	1.7	171	60	2.8	4.7	139	6.6	1
39	0	21	4	0	14.7	182	31	26.8	27.5	379	104.2	4
40	0	42	6	0	6.8	319	254	21.8	9.7	266	25.7	1
41	0	71	4	2	2	34	27	1.1	4.2	185	7.8	4
42	0	34	2	0	4.1	212	32	8.7	5.3	154	8.1	4
43	0	54	2	3	5.3	166	124	8.7	6.8	236	16	3
44	0	60	0	0	11.4	127	30	14.5	9.3	148	13.8	2
45	0	33	2	2	8.7	135	54	11.8	29	186	53.8	2
46	0	40	20	1	11.6	123	88	14.2	22	152	33.3	1
47	1	32	10	1	10.3	120	20	12.3	11.9	135	16.1	1
48	0	37	3	0	7.5	86	21	6.4	15	189	28.3	1
49	1	31	5	3	4	56	43	2.2	7.4	137	10.2	1
50	0	25	7	3	2.2	184	10	4.1	5.4	459	24.7	1
51	1	27	1	3	3.1	140	60	4.4	6.6	167	11	2
52	0	56	15	1	8.3	60	17	5	11.4	72	8.2	1
53	0	23	2	0	6	133	26	8	11.5	113	13	1
54	0	33	14	0	2.9	191	23	5.6	15.5	136	21.1	2
55	0	56	6	3	5.6	140	35	7.9	12.5	129	16.1	1
56	1	27	7	1	7.1	270	180	19.1	11.2	345	38.7	1
57	0	51	3	1	3.5	111	50	3.8	15.1	212	32	1
58	0	31	0.5	3	4.7	525	105	24.7	10.8	627	67.7	1
59	0	50	8	4	10.6	185	21	19.6	25.3	224	56.6	4
60	1	31	12	0	2	45	63	0.9	7.1	165	11.7	1
61	1	47	2	0	26.1	68	46	17.7	28	307	86	1
62	0	34	4	2	8.8	95	32	8.3	11.8	183	12.6	2
63	0	42	1	3	3.7	514	75	19.2	12.5	312	39.1	4

Table A.1: (conti.)

No.	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11	class
64	0	27	2	2	4	96	14	3.8	14.9	69	10.3	4
65	1	32	0.5	0	7.8	69	78	5.4	16.7	51	8.5	3
66	1	35	3	0	2.3	43	28	1	8.3	90	7.5	1
67	0	36	10	0	3.2	79	38	2.6	9.2	165	15.2	1
68	0	34	2	0	5.5	108	80	6	11.1	121	13.4	1
69	0	27	4	0	3.3	159	72	5.2	5	127	6.3	1
70	1	32	7	0	6.1	43	74	2.6	10.8	326	35.1	1
71	1	47	15	2	2.2	112	35	2.4	16.7	53	8.7	1
72	0	35	7	0	4.4	118	38	5.2	5.7	129	7.4	1
73	0	28	15	4	7.3	23	110	1.7	9.8	21	20.6	2
74	0	45	24	0	1.4	60	28	0.9	7.1	146	10.3	1
75	1	27	10	0	21	187	225	39.1	39.1	387	151.4	4
76	0	27	4	0	10.6	127	30	14	11	430	45.6	1
77	0	26	3	0	3.8	283	43	11	11.7	260	30.3	1
78	0	27	4	0	4.6	79	20	3.6	8.7	184	16.1	1
79	0	28	1	1	1	214	40	2.1	8.6	442	37.9	3
80	0	50	32	0	5.1	171	30	8.8	5.1	135	7	2
81	0	28	11	0	4.3	145	65	6.3	6	196	11.8	4
82	0	27	4	0	6	225	50	13.6	18.8	129	24.2	3
83	0	48	10	0	11	102	20	11.2	16.3	142	23.2	1
84	0	30	10	0	9.4	249	70	23.5	18.6	194	36.1	1
85	1	34	15	0	15.9	136	60	21.6	17.8	184	32.8	1
86	0	22	3	0	10.6	198	30	20.9	11.9	188	22.4	2
87	0	30	5	0	8.6	155	37	13.3	13.9	232	32.1	2
88	0	51	1	1	14.9	80	20	11.9	20.7	128	26.5	1
89	1	30	10	0	6.8	136	100	9.3	20.7	128	26.5	1
90	0	30	5	0	7.4	213	90	15.7	10.5	266	28	2
91	0	35	4	0	3.8	57	116	2.2	10.4	191	19.8	1
92	0	30	10	0	7.6	158	22	12	12.1	169	20.4	4
93	0	43	6	0	3.1	122	15	3.8	1.6	208	3.4	1
94	0	42	10	0	11.7	159	132	18.6	19.6	127	24.9	1
95	1	45	12	0	5.2	53	32	2.7	13.8	286	39.5	4
96	0	34	1.5	1	4.5	104	70	4.6	12.1	263	32.6	2
97	0	36	5	0	7.1	110	26	7.9	13.5	277	37.4	1
98	0	30	9	0	4.3	134	55	5.8	8.8	336	29.6	1
99	0	31	5	2	2.5	19	134	0.48	9.1	149	13.5	1
100	0	25	9	0	8.2	60	78	4.9	14.2	151	21.4	1
101	0	30	10	1	1.5	122	80	1.9	5.3	220	11.6	4
102	0	33	5	2	5.7	68	10	3.9	6.4	245	15.6	1
103	0	32	2	0	6	187	60	11.2	11	285	31.4	2
104	0	45	22	2	8.7	80	90	7	42.3	270	114.3	1
105	0	38	2	0	5.8	58	8	3.4	7.1	148	10.6	4
106	1	56	0.83	0	8.8	73	30	6.4	20	68	13.7	1
107	1	45	11	0	6.3	50	105	3.1	13.2	91	12	4
108	0	32	2	1	8.9	143	75	12.8	10.9	280	30.4	1
109	0	60	2	0	4.2	195	50	8.1	6.5	265	17.3	3
110	1	44	3	2	3.7	86	5	3.2	7.7	170	13.1	2
111	0	49	4	0	6.3	180	15	11.4	21	115	84.2	1
112	1	28	10	0	9.5	98	60	9.3	14.7	1344	19.7	1
113	0	26	2	0	8.3	82	60	6.8	26.3	330	86.9	1
114	0	39	5	0	7.5	137	14	10.3	10.7	160	17.1	1
115	0	49	9	0	3.1	150	40	4.6	7	261	18.4	1
116	0	30	1	1	17.4	76	29	13.2	24.8	229	56.7	1
117	0	52	4	1	5.7	45	27	2.6	15.4	242	37.2	1
118	0	45	3	0	5.2	67	128	3.5	11.8	230	27.1	3
119	0	53	7	0	7.4	68	30	5	8.7	140	12.2	1
120	0	29	6	0	15.7	120	40	18.8	12.3	220	27	2
121	0	28	4	0	8.9	88	28	7.8	12.3	163	20	2
122	0	38	5	2	1	128	6	1.3	5.8	145	8.4	1

B The European Barn Swallow Data Set

Table B.2: The European barn swallow data set with original values of attributes

No.	a1	a2	a3	a4	a5	a6	a7	a8	class
1	30	104	47	104	127	127	18.2	4844	2
2	30.2	107	46	109	127	126	18.3	5434	2
3	30.6	110	45	96	126	126	17.9	6498	2
4	30.6	108	46	108	126	127	18.4	5789	2
5	30.9	123	50	126	130	129	18.4	5553	2
6	30.2	112	48	100	122	122	17.7	5553	2
7	30	92	47	93	121	122	20.9	5907	1
8	30.5	99	44	99	127	126	17.9	5080	2
9	29	91	44	91	120	120	17.1	5198	1
10	29.6	102	45	104	130	129	18.1	6616	2
11	30.3	94	48	95	123	124	18.7	4962	1
12	29.6	86	47	87	120	120	17	5316	1
13	30	98	46	99	123	123	18.2	5553	1
14	30.2	90	46	89	125	125	18.7	4607	2
15	30.8	88	41	89	126	126	21.7	5198	1
16	29.5	96	46	97	125	125	18.4	4371	2
17	29.5	85	45	83	130	129	20.1	5434	2
18	30.2	85	45	86	122	122	19.5	5316	1
19	31	100	47	102	123	123	21.8	4371	1
20	30.6	129	44	129	125	125	18.2	5671	2
21	29.6	83	47	83	121	121	19.9	4607	1
22	30.1	115	47	84	123	123	19	4489	2
23	29.7	92	49	94	124	124	22.1	5080	1
24	29.2	91	44	90	124	124	20	4844	1
25	29.6	75	46	85	129	129	19.7	4253	2
26	29.2	110	44	110	123	123	16.7	4962	2
27	30	95	47	95	121	121	21.5	5789	1
28	29.4	96	44	95	123	124	17.2	4607	1
29	29.7	93	44	94	123	123	25.1	5080	1
30	30.9	105	45	105	121	120	19	5553	2
31	30.1	118	46	118	127	128	19.7	5671	2
32	30.9	84	46	86	125	126	21.2	4962	1
33	29.6	93	46	87	123	122	21.3	5316	1
34	30.1	110	43	111	126	126	18.9	5198	2
35	29.8	90	43	84	120	120	16.6	5080	1
36	30.4	89	50	91	133	133	20	5434	1
37	30.5	109	46	109	125	125	17.6	5671	2
38	29.6	108	45	109	125	125	17.1	4726	2
39	30.4	87	47	87	126	126	21.5	5198	1
40	29.6	86	46	86	124	124	21	5316	1
41	30	99	45	99	125	124	19.7	4844	2
42	30.4	90	41	97	122	122	17.2	3780	2
43	30.1	94	46	96	122	122	18.8	4607	2
44	30.2	113	47	115	126	124	19	6616	2
45	29.7	102	49	95	128	126	17.5	4726	2
46	29.6	93	44	92	122	123	21.7	5434	1
47	29.9	100	46	97	125	123	19.6	5671	1
48	29.3	97	47	96	130	130	18.5	4962	2
49	30.2	86	47	85	126	126	18	5080	1
50	29.7	85	47	85	125	125	18.8	5080	1
51	30.1	88	45	89	120	122	18.8	4489	1
52	30.4	106	48	106	124	124	18	4253	2
53	29.5	105	49	103	126	126	17.8	6025	2
54	30.2	96	48	97	126	125	18.4	5198	1
55	30	107	45	109	128	128	18.8	6498	2
56	30.9	90	49	89	129	129	20.4	5789	1
57	29.3	101	48	97	123	123	20.2	4962	1

Table B.2: (conti.)

No.	a1	a2	a3	a4	a5	a6	a7	a8	class
58	30.3	126	44	125	130	130	17.9	4607	2
59	30	88	48	88	122	121	19.2	4962	1
60	30.7	117	44	119	127	126	19.6	5316	2
61	29.5	85	48	84	118	118	17.5	5080	1
62	30.8	108	47	108	126	127	18.9	5671	2
63	30.3	97	46	95	126	125	18.4	5316	1
64	30.6	97	48	97	128	127	18.1	5434	1
65	29.6	90	48	77	122	121	18.2	5198	1
66	30.2	112	46	113	129	130	19.6	5198	2
67	29.6	93	45	91	119	118	21	5080	1
68	29.7	86	49	86	125	124	17.6	5080	1
69	28.9	74	46	103	127	126	17.3	4726	2

C The Input Data File Format for MCOCR

The file as shown in Appendix D is a sample input data file for MCOCR. An input data file represents an input data set. There are three tags in an input data file.

- **Classes**: to tell MCOCR that classes' definitions are beginning.
- **Attributes**: to tell MCOCR that attributes' definitions are beginning.
- **Objects**: to tell MCOCR that objects' details are beginning.

The order of tags "Attributes," "Classes" and "Objects" is arbitrary.

CLASS.

Format—

Classes [Class num.] [1'st class index] [2'nd class index] ... [n'th class index]

meaning:

- first field means the number of classes in the input data set
- the following fields means the class index of each class

Example—

Classes 4 c1 c2 c3 c4

The example means that there are four classes in the input data set. Their index are c1, c2, c3 and c4, respectively.

NOTE: The initial character of a class index must be 'C' or 'c', and followed by a number. □

ATTRIBUTES.

Format—

Attributes

[1'st attribute index] [num. of 1'st attribute values]

[2'nd attribute index] [num. of 2'nd attribute values]

. . . .

. . . .

[n'th attribute index] [num. of n'th attribute values]

meaning:

- each row represents an attribute
- first field means the attribute index
- second field means the number of attribute values of such attribute

Example—

Attributes

a1 3

a2 2

a3 4

a4 3

The example means that there are four attributes in the input data set. Their index are a1, a2, a3 and a4, respectively.

- Attribute a1 has three possible attribute values, i.e., 1, 2, 3.
- Attribute a2 has two possible attribute values, i.e., 1, 2.
- Attribute a3 has four possible attribute values, i.e., 1, 2, 3, 4.
- Attribute a4 has three possible attribute values, i.e., 1, 2, 3.

NOTE: The initial character of an attribute index must be 'A' or 'a', and followed by a number. □

OBJECTS.

Format—

Objects

[1'st obj idx] [1'st obj 1'st attr val] $\cdots$ [1'st obj n'th attr val] [1'st obj cls idx]

[2'st obj idx] [2'st obj 1'st attr val] $\cdots$ [2'st obj n'th attr val] [2'st obj cls idx]

. . . .

. . . .

[m'st obj idx] [m'st obj 1'st attr val] $\cdots$ [m'st obj n'th attr val] [m'st obj cls idx]

meaning:

- each row represents an object
- first field means the object index
- last field means the class index of such object
- the rest of fields mean the attribute value of each attribute

Example—

Objects

o1 3 2 4 3 c1

o2 2 2 1 2 c1

o3 1 1 2 1 c2

o4 2 2 3 3 c3

o5 3 1 4 2 c4

The example means that there are five objects in the input data set. Their index are o1, o2, o3, o4, o5.

- Object o1 belongs to class c1, and each attribute values are 3, 2, 4, 3.
- Object o2 belongs to class c1, and each attribute values are 2, 2, 1, 2.
- Object o3 belongs to class c2, and each attribute values are 1, 1, 2, 1.
- Object o4 belongs to class c3, and each attribute values are 2, 2, 3, 3.

- Object o5 belongs to class c4, and each attribute values are 3, 1, 4, 2.

NOTE: The initial character of an object index must be 'O' or 'o', and followed by a number. □

NOTE: The separator between two fields can only be SPACE.

D A Sample Input Data File for MCOCR

```
classes 4 c1 c2 c3 c4
ATTRIBUTES
a1 2
a2 2
a3 3
a4 5
a5 3
a6 3
a7 3
a8 3
a9 3
a10 3
a11 4
OBJECTS
o1 2 2 3 1 3 2 1 3 3 2 2 c1
o2 1 1 2 2 3 1 1 3 3 2 3 c1
o3 1 1 3 1 3 2 1 3 3 1 1 c1
o4 1 2 2 1 3 1 1 3 3 2 4 c1
o5 2 1 1 1 3 2 1 3 3 2 2 c3
o6 1 1 3 2 3 1 2 3 2 2 2 c1
o7 1 1 2 2 3 1 1 1 2 1 1 c1
o8 2 1 2 2 3 1 1 2 3 1 1 c1
o9 1 2 3 1 3 2 1 3 3 2 4 c1
o10 1 1 3 4 3 2 1 3 3 3 4 c2
o11 1 2 2 1 3 3 1 3 3 2 4 c1
o12 1 2 3 3 2 2 1 1 1 2 1 c2
o13 1 2 3 4 3 2 1 3 2 2 1 c2
o14 1 1 2 1 3 1 1 1 3 2 2 c1
o15 1 2 3 3 3 2 1 3 3 3 4 c1
o16 1 1 2 2 3 3 1 3 2 2 2 c1
o17 1 1 3 1 3 3 1 3 3 2 4 c1
o18 2 2 3 1 3 1 1 3 1 1 1 c4
o19 1 2 3 1 3 2 1 3 3 3 4 c1
o20 1 1 3 1 2 3 1 3 1 2 1 c2
o21 1 1 3 1 1 3 2 1 1 2 1 c1
o22 1 1 3 5 2 3 3 3 1 2 1 c1
o23 1 1 1 2 3 3 1 3 1 2 2 c1
o24 1 1 2 4 1 1 1 1 2 2 1 c3
o25 1 2 2 1 3 2 1 3 3 2 3 c1
o26 1 1 2 1 3 2 1 3 3 2 3 c1
o27 1 2 3 1 1 3 2 3 3 1 2 c1
o28 2 2 2 1 3 1 1 1 1 1 1 c1
o29 1 1 2 1 2 3 1 3 2 2 2 c1
o30 1 1 3 1 3 2 1 3 1 2 1 c1
o31 1 2 2 1 3 2 1 3 2 2 3 c1
o32 1 2 3 1 2 1 1 1 1 2 1 c1
o33 2 2 3 2 2 1 3 1 1 1 1 c1
o34 1 1 2 1 3 3 1 3 3 2 4 c1
o35 1 1 2 1 3 2 1 3 3 2 3 c1
o36 1 1 3 1 3 2 1 3 1 3 1 c1
o38 1 1 2 2 1 3 2 2 1 2 1 c1
o39 1 1 3 1 3 3 1 3 3 3 4 c4
o40 1 2 3 1 3 3 3 3 1 3 3 c1
o41 1 2 3 3 1 1 1 1 1 2 1 c4
o42 1 1 2 1 3 3 1 3 1 2 1 c4
o43 1 2 2 4 3 3 3 3 1 2 2 c3
o44 1 2 1 2 3 2 1 3 1 2 1 c2
o45 1 1 2 3 3 2 2 3 3 2 4 c2
o46 1 2 3 2 3 2 2 3 3 2 3 c1
o47 2 1 3 2 3 2 1 3 2 2 2 c1
o48 1 2 2 1 3 2 1 3 2 2 3 c1
o49 2 1 3 4 2 1 1 2 1 2 1 c1
o50 1 1 3 4 2 3 1 3 1 3 2 c1
```