Chapter 5 Illustration

5.1 Discretization

Five data sets are demonstrated to present the effectiveness of the proposed extended Chi2 algorithm. The five data sets are taken from the University of California, Irvine's repository of machine learning databases (http:// www.ics.uci.edu/~mlearn/MLSummary.html).

5.1.1 The Data

The five data sets used in the experiment are the Bupa Liver Disorders, the Glass Types, the Heart Disease, the Iris Plants, and the Breast Cancer. They have different types of attributes. The Bupa Liver Disorders data, Glass Types data, and the Iris Plants data are of the type with continuous attributes, the Breast Cancer data are of ordinal discrete ones, while the Heart Disease data shows mixed attributes (numeric and discrete). The five data sets are described below:

(1) The Bupa Liver Disorders Data

This data set contains 345 instances (145 instances that are normal; 200 instances of a liver malfunction), where each instance is described using six numeric attributes: MCV, ALKPHOS, SGPT, SGOT, GAMMAGT, and DRINKS.

(2) The Glass Types Date

This data set contains 214 instances (70 instances of building windows that are float processed; 76 instances of building windows that are non-float processed; 17 instances of vehicle windows float processed; 13 instances of containers; 9

instances of tableware; 29 instances of headlamps), each instance is described using nine numeric attributes: RI, NA, MG, AL, SI, K, CA, BA, and FE.

(3) The Iris Plants Data

This data set contains 150 instances (50 instances of setosa; 50 instances of versicolor; 50 instances of verginica); each instance is described using four numeric attributes: sepal-length, sepal-width, petal-length, and petal width.

(4) The Breast Cancer Data

This data set contains 699 instances, where 16 instances have missing attributes values. Removing instances with missing attributes values, we use 683 instances (444 instances of benign; 239 instances of malignant), where each instance is described using nine attributes: clump thickness, uniformity of cell size, uniformity of cell shape, marginal adhesion, single epithelial cell size, bare nuclei, bland chromatin, normal nucleoli, and mitoses.

(5) The Heart Disease Data

This data set contains 297 instances (160 instances of 0; 54 instances of 1; 35 instances of 2; 35 instances of 3; 13 instances of 4), where each instance is described using eight nominal attributes: SEX, CP, FBS, RESTECG, EXANG, SLOPE, CA, and THALPUL; and five numeric attributes: AGE, TRESTBPS, CHOL, THALACH, and OLDPEAK.

5.1.2 Experimental Results

We ran See5 on both the original data sets and the discretized data sets. The parameters of See5 utilize its default setting. The ten-fold cross-validation test method is applied to all data sets. The data set is divided into 10 parts of which nine parts are used as training sets and the remaining one part as the testing set. The

experiments were repeated 10 times. The final predictive accuracy is taken as the average of the 10 predictive accuracy values.

The extended Chi2 algorithm is compared with the original Chi2 algorithm and modified Chi2 algorithm with the predefined inconsistency rate (δ) value equal to 0 in the experiment. The experimental process includes two steps:

Step 1: Discretization:

All five data sets are discretized using the original Chi2 algorithm, the modified Chi2 algorithm, the extended Chi2 algorithm, and Boolean Reasoning algorithm.

Step 2: Comparison:

The discretized data sets are sent into See5. The predictive accuracy and its standard deviation of these methods are listed in Table 5.1. From Table 5.1, we know that the predictive accuracy of the extended Chi2 algorithm outperforms other discretization algorithms.

Table 5.1 The Predictive Accuracy Using See5 With the Discretization Algorithm

Data Set	See5				
	Continuous	Original Chi2 Algorithm	Modified Chi2 Algorithm	Extended Chi2 Algorithm	Boolean Reasoning Algorithm
Bupa	67.5 ± 2.4%	65.2 ± 3.2%	67.5 ± 1.9%	68.4 ± 2.7%	68.1 ± 2.3%
Glass	68.6 ± 2.5%	93.1 ± 2.1%	93.4 ± 2.3%	93.5 ± 1.3%	71.9 ± 2.8%
Iris	94.0 ± 2.1%	94.0 ± 2.1%	93.3 ± 2.2%	94.0 ± 2.1%	96.0 ± 1.8%
Breast Cancer	94.9 ± 0.8%	95.5 ± 1.0%	96.0 ± 0.9%	96.5 ± 0.8%	95.2 ± 0.8%
Heart dissease	51.9 ± 1.4%	52.5 ± 2.3%	53.2 ± 2.7%	54.2 ± 1.7%	55.9 ± 2.6%

The tree sizes using See5 with different discretization methods shown in Table 5.2. From Table 5.2, we know that although the extended Chi2 algorithm has no significant difference in tree size compared to the original and modified Chi2 algorithms, it is in fact significantly smaller than when using the original data with See5.

Table 5.2 The Tree Size Comparison of the Five Methods

Data Set	See5				
	Continuous	Original Chi2 Algorithm	Modified Chi2 Algorithm	Extended Chi2 Algorithm	Boolean Reasoning Algorithm
Bupa	27.1 ± 1.7	15.9 ± 1.0	12.7 ± 0.6	9.9 ± 0.6	30.3 ± 2.1
Glass	24.0 ± 0.7	9.9 ± 0.1	9.8 ± 0.1	9.2 ± 0.2	23.5 ± 0.9
Iris	4.6 ± 0.2	3.7 ± 0.2	3.0 ± 0.0	3.0 ± 0.0	3.9 ± 0.1
Breast Cancer	10.3 ± 0.9	8.6 ± 0.7	8.8 ± 0.8	9.1 ± 0.8	9.2 ± 0.3
Heart dissease	46.0 ± 0.9	34.8 ± 1.7	34.1 ± 0.7	36.0 ± 1.1	34.7 ± 1.2

5.2 The β-reducts

In this section, a simple example is used to illustrate the proposed procedure shown in section 4.2. Also, a medical case is demonstrated to present the effectiveness of our proposed approach.

5.2.1 A Simple Example

The data sets taken from the literature (Beynon, 2001) are given in Table 5.3. There exists a set of objects $U(1,2,...,7)$ contained in the rows of the table, with the columns denoting the condition attributes $C(a,b,c,d,e,f)$ of these objects, and a related decision attribute D.

Table 5.3 Information System

Objects	a	b	c	d	e	f	D
1	1	1	1	1	1	1	M
2	1	0	1	0	1	1	M
3	0	0	1	1	0	0	M
4	1	1	1	0	0	1	F
5	1	0	1	0	1	1	F
6	0	0	0	1	1	0	F
7	1	0	1	0	1	1	F

In this information system the objects have been classified into one of two categories, M and F. In the information system, objects 1, 3, 4, and 6 are unambiguously classified, in the sense that all objects with a given set of attribute values are assigned to the same category. Objects 2, 5, and 7 are ambiguously classified since they have the same combination of condition attributes, but they are not all classified to the same decision category. Subsequently, the condition classes of objects as groupings of indiscernible objects are:

$C^* = \{C_1, C_2, C_3, C_4, C_5\}$, where $C_1 = \{1\}, C_2 = \{2,5,7\}, C_3 = \{3\}, C_4 = \{4\}, C_5 = \{6\}$.

Similarly, the decision classes of the categories are:

$$D^* = \{M, F\}, \quad where \quad D_M = \{1,2,3\} \text{ and } D_F = \{4,5,6,7\}.$$

Choose the precision parameter value of the information system

Based on (3.1), $\xi(C,D) = \max(m_1, m_2)$, where

$m_1 = 1 - \min[0.5 < c(C,D)], m_2 = \max[c(C,D) < 0.5]$

$c(C_1, D_M) = 1 - \frac{1}{1} = 0$ $\quad$ $c(C_1, D_F) = 1 - \frac{0}{1} = 1$

$c(C_2, D_M) = 1 - \frac{1}{3} = \frac{2}{3}$ $\quad$ $c(C_2, D_F) = 1 - \frac{2}{3} = \frac{1}{3}$

$c(C_3, D_M) = 1 - \frac{1}{1} = 0$ $\quad$ $c(C_3, D_F) = 1 - \frac{0}{1} = 1$

$c(C_4, D_M) = 1 - \frac{0}{1} = 1$ $\quad$ $c(C_4, D_F) = 1 - \frac{1}{1} = 0$

$c(C_5, D_M) = 1 - \frac{0}{1} = 1$ $\quad$ $c(C_5, D_F) = 1 - \frac{1}{1} = 0$

$m_1 = 1 - \min(\frac{2}{3}, 1) = \frac{1}{3}$

$m_2 = \max(0, \frac{1}{3}) = \frac{1}{3}$.

Thus, $\xi(C,D) = \max(\frac{1}{3}, \frac{1}{3}) = \frac{1}{3}$.

Therefore, the precision parameter value is equal to $\frac{1}{3}$.

Find the full set of β-reducts

Since the β value is equal to $1/3$, then $\underline{C}_{\beta}(D_M) = C_1 \cup C_3 = \{1,3\}$ and $\underline{C}_{\beta}(D_F) = C_2 \cup C_4 \cup C_5 = \{2,4,5,6,7\}$. The discerniblity matrix, *M* (*S*), for the seven elementary sets presented, is shown in Table 5.4. The relative discernibility functions are:

$$\Delta_{\underline{\beta}}(1) = (d+e)(b+d)(a+b+c+f)$$

$$= be + ad + bd + cd + df$$

$$\Delta_{\underline{\beta}}(2) = (b+e)(a+c+d+f)$$

$$= ab + bc + bd + bf + ae + ce + de + ef$$

$$\Delta_{\underline{\beta}}(3) = (a+b+d+f)(c+e)(a+d+e+f)$$

$$= ac + ae + cd + de + be + cf + ef$$

$$\Delta_{\underline{\beta}}(4) = (d+e)(b+e)(a+b+d+f)$$

$$= ae + bd + be + ed + ef$$

$$\Delta_{\underline{\beta}}(5) = (b+d)(a+d+e+f)$$

$$= d + ab + be + bf$$

$$\Delta_{\underline{\beta}}(6) = (c+e)(a+b+c+f)(a+c+d+f)$$

$$= c + ae + ef + bed$$

$$\Delta_{\underline{\beta}}(7) = (b+d)(a+d+e+f)$$

$$= \Delta_{\underline{\beta}}(5)$$

$$\Delta_{\underline{\beta}}(D_M) = \Delta_{\underline{\beta}}(1) * \Delta_{\underline{\beta}}(3)$$

$$= (be + ad + bd + cd + df)(ac + ae + cd + de + be + cf + ef)$$

$$= cd + be + ade + def$$

$$\Delta_{\underline{\beta}}(\mathrm{D_F}) = \Delta_{\underline{\beta}}(2) * \Delta_{\underline{\beta}}(4) * \Delta_{\underline{\beta}}(5) * \Delta_{\underline{\beta}}(6) * \Delta_{\underline{\beta}}(7)$$

$$= (ab + bc + bd + bf + ae + cd + de + ef)(ae + bd + be + ed + ef)$$
$$(d + ab + be + bf)(c + ae + ef + bde)(d + ab + be + bf)$$

$$= abe + bce + cde + abe + def + bde + bcd + bcf + bef$$

$$\Delta_{\underline{\beta}}(D) = \Delta_{\underline{\beta}}(D_M) * \Delta_{\underline{\beta}}(D_F)$$

$$= (cd + be + ade + def)(abe + bce + cde + abe + def + bde + bcd + bcf + bef)$$

$$= abe + bce + cde + ade + def + bde + bcd + bef\ .$$

By step 1, we have eight subsets, which are:

$\{a,b,e\},\{b,c,e\},\{c,d,e\},\{a,d,e\},\{d,e,f\},\{b,d,e\},\{b,c,d\}$ and $\{b,e,f\}$.

Since $\{c,d\}$ has the least number of attributes, it is selected. In Table 5.5, the M (S)-information system for β-reduct $\{c,d\}$ is presented. By step 2, we have four β-reducts, which are: $\{c,d\},\{b,e\},\{d,e,f\}$ and $\{a,d,e\}$. Since β-reducts $\{c,d\}$ has the least number of combinations of values of its attributes, it is selected for further study. In Table 5.5, the $M(S)$-information system for β-reducts $\{c,d\}$ is presented.

Table 5.4 Discernibility Matrix

	1	2	3	4	5	6	7
1		—	—	*d,e*	*b,d*	*a,b,c,f*	*b,d*
2	—		—	*b,e*	—	*a,c,d,f*	—
3	—	—		*a,b,d,f*	*a,d,e,f*	*c,e*	*a,d,e,f*
4	*d,e*	*b,e*	*a,b,d,f*		—	—	—
5	*b,d*	—	*a,d,e,f*	—		—	—
6	*a,b,c,f*	*a,c,d,f*	*c,e*	—	—		—
7	*b,d*	—	*a,d,e,f*	—	—	—	

Table 5.5 *M (S)*-Information System

Objects	*c*	*d*	*D*
1	1	1	M
2	1	0	M
3	1	1	M
4	1	0	F
5	1	0	F
6	0	1	F
7	1	0	F

We also are interested in the elimination of superfluous values of condition attributes in the *M* (*S*)-information system. To do this, we must compute the relative attributes values of subset $\{c,d\}$, based on the *M* (*S*)-discerniblity matrix constructed for the *M (S)*-information system. Table 5.6 presents the *M* (*S*)-discerniblity matrix for the β-reducts $\{c,d\}$. The relative discernibility functions are:

$f_1(A) = cd$

$f_2(A) = cd$

$f_3(A) = cd$

$f_4(A) = d$

$f_5(A) = d$

$f_6(A) = c$

$f_7(A) = d$

Table 5.7 shows the information system's final version in the subspace $\{c,d\}$.

Table 5.6 *M (S)*-Discernibility Matrix

	1	2	3	4	5	6	7
1		—	—	*d*	*d*	*c*	*d*
2	—		—	—	—	*c,d*	—
3	—	—		*d*	*d*	*c*	*d*
4	*d*	—	*d*		—	—	—
5	*d*	—	*d*	—		—	—
6	*c*	*c,d*	*c*	—	—		—
7	*d*	—	*d*	—	—	—	

Table 5.7 Final Version of Information System (literature data)

Objects	*c*	*d*	*D*
1	1	1	M
2	1	0	M
3	1	1	M
4	*	0	F
5	*	0	F
6	0	*	F
7	*	0	F

Note: " * " indicates: don't care.

Rules extraction

According to Table 5.7, the generalized rules are listed in Table 5.8. Comparing the implementation results from the proposed method with the literature approach (the extracted rules and results are listed in Table 5.9), the extracted rules' numbers generated by our method are less than those of the literature's approach, though the classification accuracy in the literature's approach is as good as our proposed method.

Table 5.8 Decision Rules (by the proposed approach)

Rules	Accuracy
1. If *c*=1 and *d*=1 then *D*=M	100% (2/2)
2. If *d*=0 then *D*=F	75% (3/4)
3. If *c*=0 then *D*=F	100% (1/1)

Notes: (/) indicates (number of correct instances/number of total instances).

Table 5.9 Decision Rules (by the Beynon approach)

Rules	Accuracy
1. If *b*=1 and *e*=1 then *D*=M	100% (1/1)
2. If *b*=0 and *e*=1 then *D*=F	75% (3/4)
3. If *b*=0 and *e*=0 then *D*=M	100% (1/1)
4. If *b*=1 and *e*=0 then *D*=F	100% (1/1)

Notes: (/) indicates (number of correct instances/number of total instances).

5.2.2 A Medical Case

This case utilizes medical data to diagnose liver malfunctions. The data are from the general medical examination items at a hospital located in Taipei, Taiwan. The

examination data has fifteen items. They are: Age, Neutrophil, Lymphocyte, Moncyte, Basophil, GLUAC, ALK-P, GOT, GPT, γ-GT, D-Bil., T-Protein, TG, BUN, and Uric Acid. These items are characterized by multi-dimensional information about the current health status of patients, which makes it difficult to diagnose other diseases based on such a large amount of information. Until now, the relationship between the medical examination data and liver malfunction symptom is still ambiguous.

In this case 168 instances are collected. These instances are separated into a training set that includes 101 instances (54 instances that are normal; 47 instances of liver malfunctions) and a test set that includes 67 instances (35 instances that are normal; 32 instances of liver malfunctions). Labeling "liver malfunction patients" is based on the medical history of the patients as judged by medical doctors.

Using the proposed approach

Since VPRS needs the data in a categorical form, the continuous attributes must be discretized before the VPRS analysis is performed. In this case the items of the medical examination standard (MES) are utilized to discretize the continuous attributes. The results are listed in Table 5.10. From Table 5.10, we know that each condition attribute is classified into two or three ranges.

Table 5.10 Condition Attributes Ranges for MES Discretization

Examination Items	Range '1'	Range '2'	Range '3'
Age (a_1)	23~34	35~55	56~63
Neutrophil (a_2)	0.0~36.9	37.0~75.0	75.1~
Lymphovyte (a_3)	0.0~19.9	20.0~55.0	55.1~
Moncyte (a_4)	0.0~2.4	2.5~10.0	10.1~
Basophil (a_5)	—	0.0~2.0	2.1~
GLUAC (a_6)	0~69	70~110	111~
ALK-P (a_7)	0~59	60~205	206~
GOT (a_8)	0~7	8~35	36~
GPT (a_9)	—	0~35	36~
γ -GT (a_{10})	—	0~45	46~
D-Bil. (a_{11})	0	0.1~0.5	0.6~
T-Protein (a_{12})	0.0~6.2	6.3~8.5	8.6~
TG (a_{13})	0~59	60~105	106~
BUN (a_{14})	0~7	8~25	26~
Uric Acid (a_{15})	0.0~2.4	2.5~8.0	8.1~

In this information system the objects have been classified into one of two categories, 0 (normal) and 1 (malfunction). The condition classes of objects can distinguish 40 groups, and the precision parameter value is equal to $1/6$. Following the method of analysis given previously, four subsets and a β-reducts can be obtained. The subsets are: $\{a_2,a_7,a_8,a_9,a_{10},a_{13}\}$, $\{a_3,a_7,a_8,a_9,a_{10},a_{13}\}$, $\{a_7,a_8,a_9,a_{10},a_{13},a_{15}\}$, and $\{a_7,a_8,a_9,a_{10},a_{12},a_{13}\}$; the β-reduct is $\{a_7,a_8,a_9,a_{10},a_{13}\}$. The final version of the information system in the subspace $\{a_7,a_8,a_9,a_{10},a_{13}\}$ is shown in Table 5.11.

Table 5.11 Final Version of the Information System (liver data)

Objects	a_7	a_8	a_9	a_{10}	a_{13}	D
1	2	2	2	2	2	0
2	3	*	2	*	3	0
3	*	*	3	*	*	1
4	2	*	*	*	3	1
5	*	3	*	*	*	1
6	*	*	3	*	*	1
7	*	*	3	*	*	1
8	*	3	*	*	*	1
9	*	*	*	3	*	1
10	*	*	*	3	*	1
11	*	*	*	3	*	1
12	3	*	*	*	2	1
13	*	3	*	*	*	1
14	*	*	*	3	*	1
15	*	*	3	*	*	1
16	*	*	3	*	*	1
17	1	*	*	*	*	1

Note: " * " indicates: don't care.

According to Table 5.11, the extracted rules are listed in Table 5.12. From Table 5.12, we know that the instances of the test set at rule 2 and rule7 are null, while rule 8 shows only one instance in the test set. Since these rules are not a matter for the judgment of liver diseases, they are deleted. The final extraction rules are listed in Table 5.13.

Table 5.12 Results of Rule Extraction (VPRS)

Rules	Accuracy (%)	
	Training Set	Test Set
1. If $60 \leq$ ALK-P ≤ 205, $8 \leq$ GOT ≤ 35, $0 \leq$ GPT ≤ 35, $0 \leq \gamma$ -GT ≤ 45 and $60 \leq$ TG ≤ 105, then one is normal.	100% (52/52)	94.59% (35/37)
2. If $206 \leq$ ALK-P, $0 \leq$ GPT ≤ 35 and $106 \leq$ TG, then one is normal.	100% (1/1)	—
3. If $36 \leq$ GPT, then one has a malfunction.	96.43% (27/28)	100% (13/13)
4. If $60 \leq$ ALK-P ≤ 205 and $106 \leq$ TG, then one has a malfunction.	100% (4/4)	100% (2/2)
5. If $36 \leq$ GOT, then one has a malfunction.	100% (8/8)	100% (8/8)
6. If $46 \leq \gamma$ -GT, then one has a malfunction.	100% (6/6)	100% (6/6)
7. If $206 \leq$ ALK-P and $60 \leq$ TG ≤ 105, then one has a malfunction.	100% (1/1)	—
8. If ALK-P ≤ 59, then one has a malfunction.	100% (1/1)	100% (1/1)

Notes: 1. "—" indicates the instance in the set is null.

2. (/) indicates (number of correct instances/number of total instances).

Table 5.13 Final Results of Rule Extraction (VPRS)

Rules	Accuracy (%)	
	Training Set	Test Set
1. If $60 \leq$ ALK-P ≤ 205, $8 \leq$ GOT ≤ 35, $0 \leq$ GPT ≤ 35, $0 \leq \gamma$ -GT ≤ 45 and $60 \leq$ TG ≤ 105, then one is normal.	100% (52/52)	94.59% (35/37)
2. If $36 \leq$ GPT, then one has a malfunction.	96.43% (27/28)	100% (13/13)
3. If $60 \leq$ ALK-P ≤ 205 and $106 \leq$ TG, then one has a malfunction.	100% (4/4)	100% (2/2)
4. If $36 \leq$ GOT, then one has a malfunction.	100% (8/8)	100% (8/8)
5. If $46 \leq \gamma$ -GT, then one has a malfunction.	100% (6/6)	100% (6/6)

Notes: (/) indicates (number of correct instances/number of total instances).

Using the neural network approach

In this section the *Professional II Plus* software package (Neural Ware, Inc., 1992) is used to perform the computation in order to obtain the structure with a maximum classification rate. After trial and error, we choose 0.25 and 0.30 as the learning rates in the hidden layer and the output layer, respectively. The momentum is set at 0.95, and the number of iterations is set at 20000. Structure 15-10-1 is the optimal structure through the trained back-propagation neural network.

After the features selection, eight features are deleted and seven features are retained. They are Lymphocyte, Monocyte, GOT, GPT, γ-GT, D-Bil, and TG. These seven features are used to retrain a new network. Structure 7-5-1 is chosen for further analysis.

After pruning the unnecessary connections from network 7-5-1, only three attributes, GOT, GPT, and γ-GT could affect the result. The simplified 7-5-1 structure is used to extract the rules, and the results are listed in Table 5.14.

Table 5.14 Results of Rule Extraction (neural networks)

Rules	Accuracy (%)	
	Training Set	Test Set
1. If GOT $\leq$ 35,GPT $\leq$ 35 and γ-GT $\leq$ 45, then one is normal.	88.14% (52/57)	90% (36/40)
2. If 36 $\leq$ GOP, then one has a malfunction.	100% (12/12)	100% (9/9)
3. If 36 $\leq$ GPT, then one has a malfunction.	97.56% (40/41)	100% (23/23)
4. If 46 $\leq$ γ-GT, then one has a malfunction.	100% (6/6)	100% (11/11)

Notes: (/) indicates (number of correct instances/number of total instances).

A comparison

A comparison is made between the proposed method and the neural networks. The results are shown in Table 5.15. Not only is the accuracy of the performance by the neural networks lower than those of our proposed method, but the results obtained by the neural networks also have less meaning from a medical geography point of view.

Table 5.15 Comparison of the Implementation Results

Method	Extraction Rules	Accuracy	
		Training Set	Test Set
VPRS	1. If $60 \le$ ALK-P ≤ 205, $8 \le$ GOT ≤ 35, $0 \le$ GPT ≤ 35, $0 \le \gamma$ -GT ≤ 45 and $60 \le$ TG ≤ 105, then one is normal.	100% (52/52)	94.59% (35/37)
	2. If $36 \le$ GPT, then one has a malfunction.	96.43% (27/28)	100% (13/13)
	3. If $60 \le$ ALK-P ≤ 205 and $106 \le$ TG, then one has a malfunction.	100% (4/4)	100% (2/2)
	4. If $36 \le$ GOT, then one has a malfunction.	100% (8/8)	100% (8/8)
	5. If $46 \le \gamma$ -GT, then one has a malfunction.	100% (6/6)	100% (6/6)
Neural Network	1. If GOT ≤ 35, GPT ≤ 35 and γ -GT ≤ 45, then one is normal.	88.14% (52/57)	90% (36/40)
	2. If $36 \le$ GOP, then one has a malfunction.	100% (12/12)	100% (9/9)
	3. If $36 \le$ GPT, then one has a malfunction.	97.56% (40/41)	100% (23/23)
	4. If $46 \le \gamma$ -GT, then one has a malfunction.	100% (6/6)	100% (11/11)

Notes: (/) indicates (number of correct instances/number of total instances).