Preliminary Analysis for Dataset #3
Statistic and Syntactic

Henry Xiao

xiao@cs.queensu.ca

School of Computing

Queen's University

Technique Preamble

Two mining techniques from two different aspects - *Bayes Methods* vs. *Rule Base Systems*.

- Bayes Method - a *statistical* approach.

- Rule Base System - a *syntactical* approach.

- Mathematical background and computational complexity.

- Structural description and applied model.

- Relations between the two techniques.

Rule Base System

We mainly look at two methods here - *DecisionTable* and *PRISM*.

- Both methods are based on rule discovery and application.

- *DecisionTable* is more or less DecisionTree like.

- *PRISM* is "if - then" structured.

- *DecisionTable* uses the correlation to select an attribute set.

- *PRISM* uses the information gain to get conditions.

- In general, *PRISM*'s rule building mechanism allows more rules.

- i.e. $a \wedge b \Longrightarrow 3$ and $c \wedge d \Longrightarrow 3$.

Rule Example

Two segments of the rules from both methods are showed.

● *DecisionTable*:

```
Rules:
================================================================================
attr_1              attr_4           attr_7            attr_10           class
================================================================================
'(0.4485-0.5535]'   '(5.3715-inf)'   '(6.4775-inf)'    '(1.8255-inf)'    3
'(0.5535-2.768]'    '(5.3715-inf)'   '(6.4775-inf)'    '(1.8255-inf)'    3
'(-0.0025-0.4485]'  '(5.3715-inf)'   '(6.4775-inf)'    '(1.8255-inf)'    3
'(-inf--0.464]'     '(5.3715-inf)'   '(6.4775-inf)'    '(1.8255-inf)'    3
'(-0.464--0.0025]'  '(5.3715-inf)'   '(6.4775-inf)'    '(1.8255-inf)'    3
'(-inf--0.464]'     '(4.7045-5.3715]' '(6.4775-inf)'   '(1.8255-inf)'    3
'(0.4485-0.5535]'   '(5.3715-inf)'   '(5.827-6.4775]'  '(1.8255-inf)'    3
'(0.5535-2.768]'    '(5.3715-inf)'   '(5.827-6.4775]'  '(1.8255-inf)'    3
'(-inf--0.464]'     '(4.24-4.7045]'  '(6.4775-inf)'    '(1.8255-inf)'    3
'(-0.464--0.0025]'  '(5.3715-inf)'   '(5.827-6.4775]'  '(1.8255-inf)'    3
'(-0.0025-0.4485]'  '(5.3715-inf)'   '(5.827-6.4775]'  '(1.8255-inf)'    3
'(-inf--0.464]'     '(5.3715-inf)'   '(5.827-6.4775]'  '(1.8255-inf)'    3
'(0.5535-2.768]'    '(5.3715-inf)'   '(6.4775-inf)'    '(1.723-1.8255]'  2
```

● *PRISM*:

```
Prism rules
----------
If attr_7 = '(-inf-3.047]' then 1
If attr_9 = '(1.753-3.1645]' then 2
If attr_10 = '(0.5915-1.1355]'
   and attr_2 = '(3.8965-inf)' then 2
If attr_10 = '(0.5915-1.1355]'
   and attr_4 = '(4.7045-5.3715]' then 2
If attr_10 = '(0.5915-1.1355]'
   and attr_8 = '(0.9835-2.0725]' then 2
If attr_10 = '(0.5915-1.1355]'
   and attr_6 = '(3.9855-4.5405]' then 2
If attr_10 = '(0.5915-1.1355]'
   and attr_5 = '(1.054-1.3095]' then 2
If attr_10 = '(0.5915-1.1355]'
```
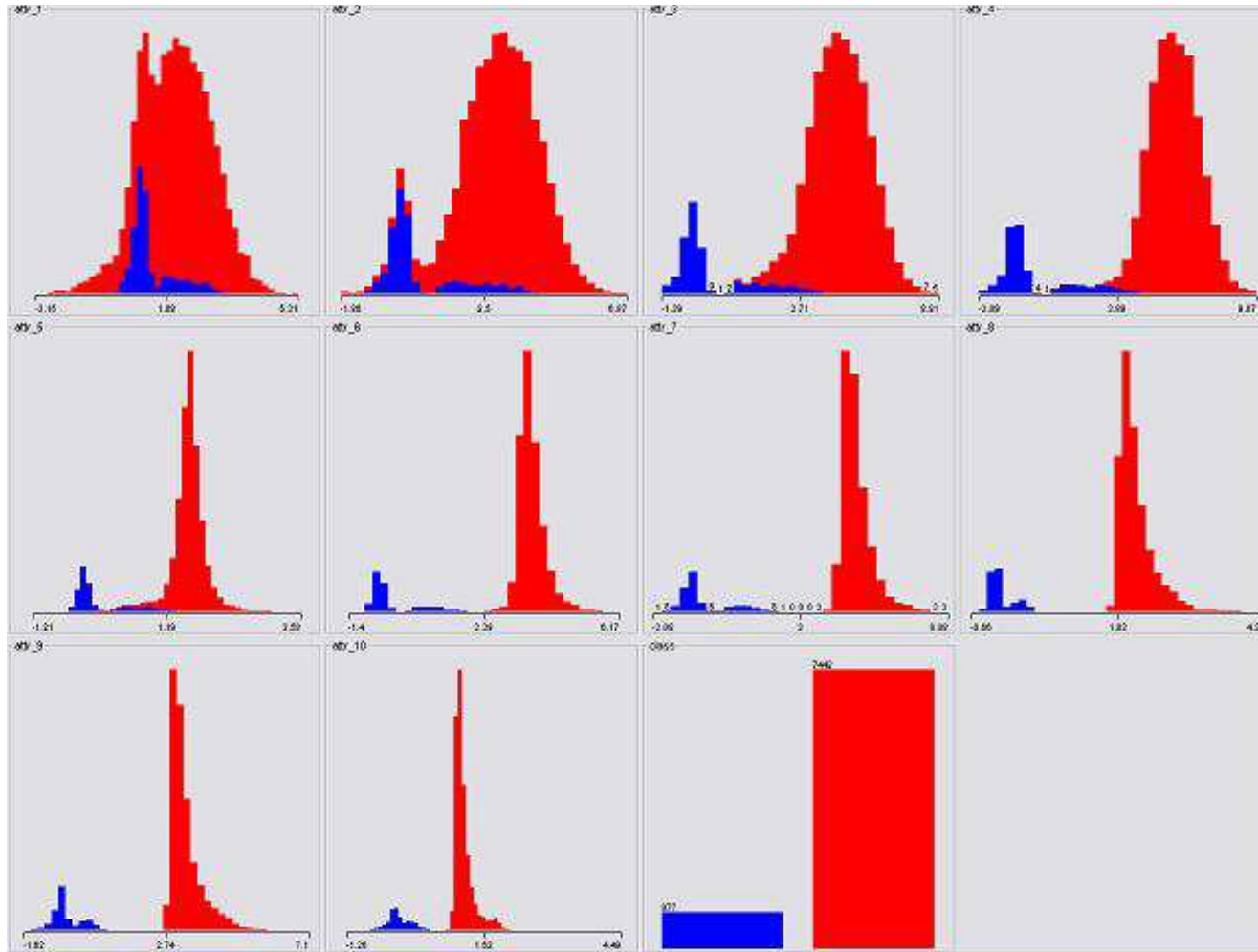
Dataset Preliminary

Dataset 3 preprocessed:

- Add a top row for attribute description.

- Name the last column as "class".

- Change the last column values to make a 2-class case.(i.e. value 8 to 1, and value 4 and 7 to 2.)

- Change the last column values to make a 3-class case.(i.e. value 8 to 1, value 7 to 2, and value 4 to 3.)
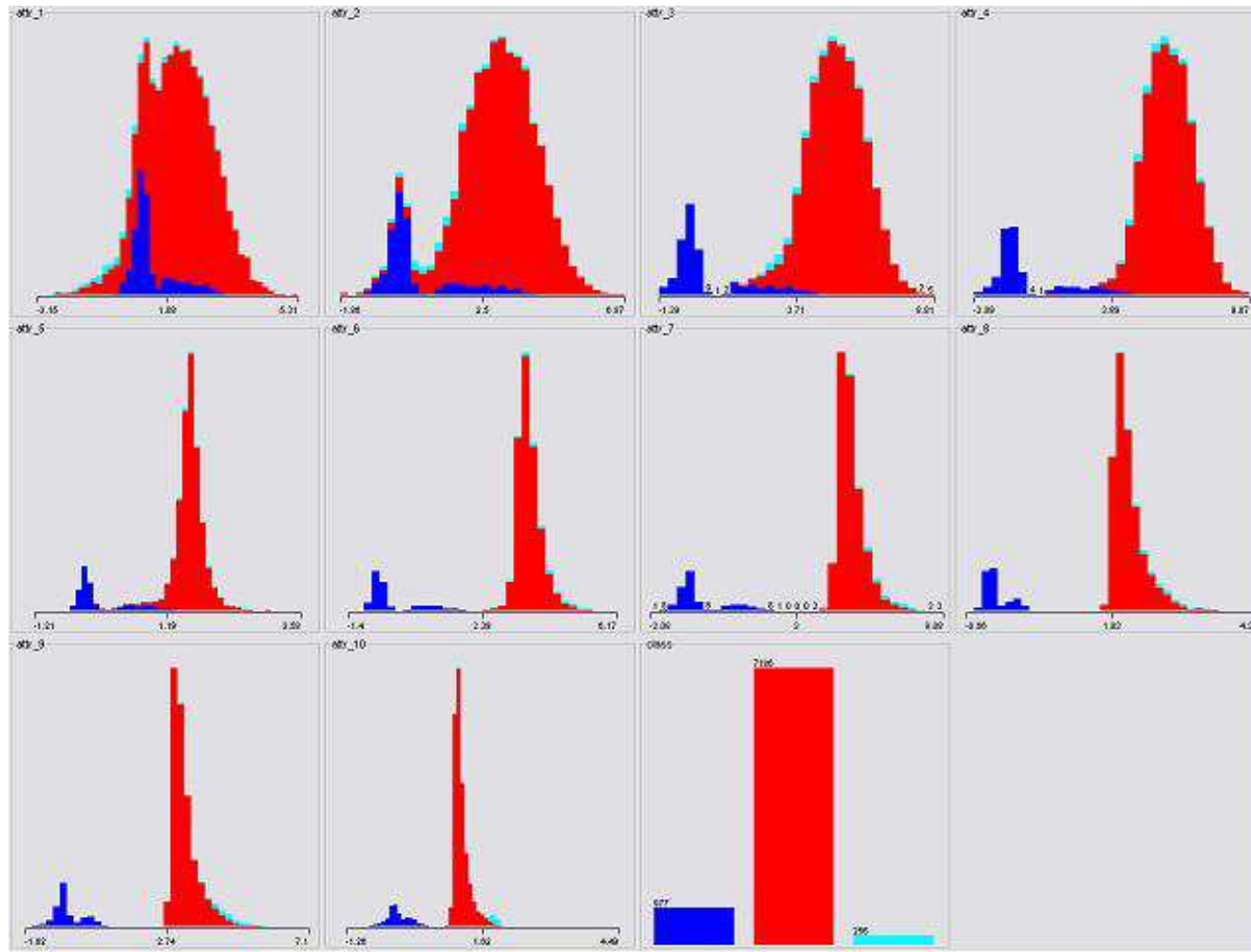
2-Class Visualization

The case is trivial. (Attribute 7, 8, 9, 10)

3-Class Visualization

Difficult to discriminate class 2 and 3.

Preliminary Result

The results of applying 3 methods with all attributes∗ are listed below.

\	BayesNet(95.2497)			DecisionTable(98.9521)			PRISM(98.7426		
Class	1	2	3	1	2	3	1	2	3
1	330	0	0	330	0	0	330	0	0
2	0	2328	126	0	2438	16	0	2437	11
3	0	10	69	0	14	65	0	17	60

Result Cont's

Remarks on the preliminary results.

- *DecisionTable* uses an attribute set of $\{1, 4, 7, 10\}$.

- The information gain ranking is: 10,9,7,8,6,4,5,3,2,1

- *DecisionTable* uses the correlation to select an attribute set.

- 10 attributes listed in the Bayes network.

- *DecisionTable* and *PRISM* difference.

Discussion and Future Plan

Two techniques are interesting to be compared here.

- *BayesNet* outperformed by *DecisionTable* and *PRISM* with all attributes.

- Two rule base methods rely on attribute's statistical information.

- With attribute selection?

- Attribute 10!

Method	BayesNet	DecisionTable	PRISM
Correctness	98.7076	98.7076	97.2407

PRISM classifies all class "3" to "2"!

Ending

Questions regarding Analysis results?

Information Site: http://www.cs.queensu.ca/home/xiao/dm.html

E-mail: xiao@cs.queens.ca

Thank you