Result of Mining Dataset 1
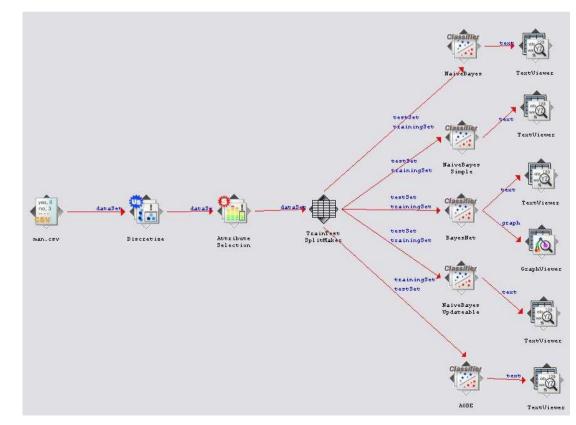## *Final Report*

Henry Xiao

`xiao@cs.queensu.ca`

School of Computing

Queen's University

Mining Process

We use *Weka* to analysis the dataset with different Bayesian Approaches. The following model is contructed using *Weka Knowledge Flow*.

Bayes Net

We mainly use the *Bayes Net* in our final mining for Dataset 1.

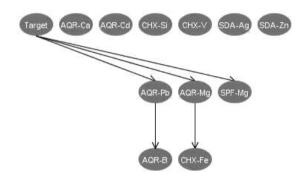- Use 12 attributes from InfoGainArributeEval.

- Use 11 attributes from ClassifierSubsetEval(*Naive set*).

- Use 12 attributes from ClassifierSubsetEval(*BayesNet set*).

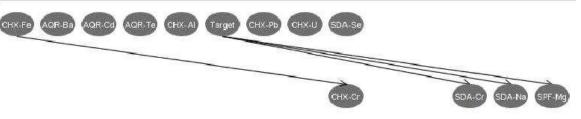- Combine the "useful" attributes to the final set.

Combined Attribute Set

The Bayes Networks for the Naive set and the BayesNet set are showed in below two figures.
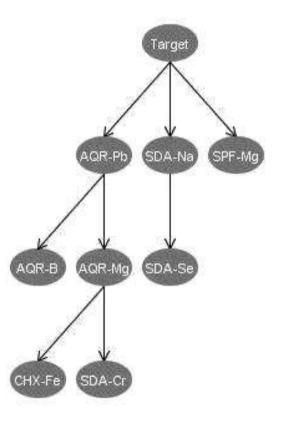
- Naive set network:

- BayesNet set network:

Combined Attribute Set

We combined the two attribute sets to a new attribute set. The network for the new set is demonstrated in the figure.

Mining Result

We use the combined attribute set and 66% training-testing split for the final mining.

■ Confusion Matrix:
$$\begin{pmatrix} a & b & classified\ as \\ 38 & 0 & (-\infty \sim 0.1] \\ 8 & 1 & (0.9 \sim \infty) \end{pmatrix}$$
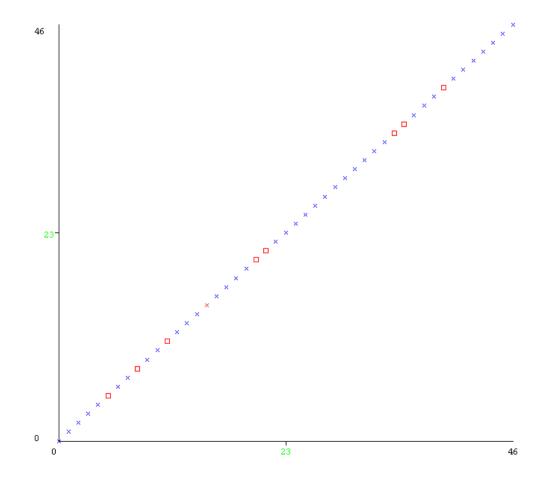
■ Total number of testing instance: 47.

■ Correctly Classified Instances: 39 - 82.9787%.

■ Time taken to build model: 0.02 seconds.

Error Display

The wrong classified instances are marked in red at the following chart.

Discussion

Some issues we encountered here.

- Attribute Selection - are the correlations useful for Bayes Net.

- Testing Strategy and Confidence Level- Bootstrap and Cross Validation.

- Improvement - Geometric information (AMGN, AMGE).

- Relation with other techniques.

Ending

Questions regarding Analysis results?

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

E-mail: xiao@cs.queens.ca

Thank you