



Analysis of Dataset 2

Final Report

Henry Xiao

xiao@cs.queensu.ca

School of Computing
Queen's University



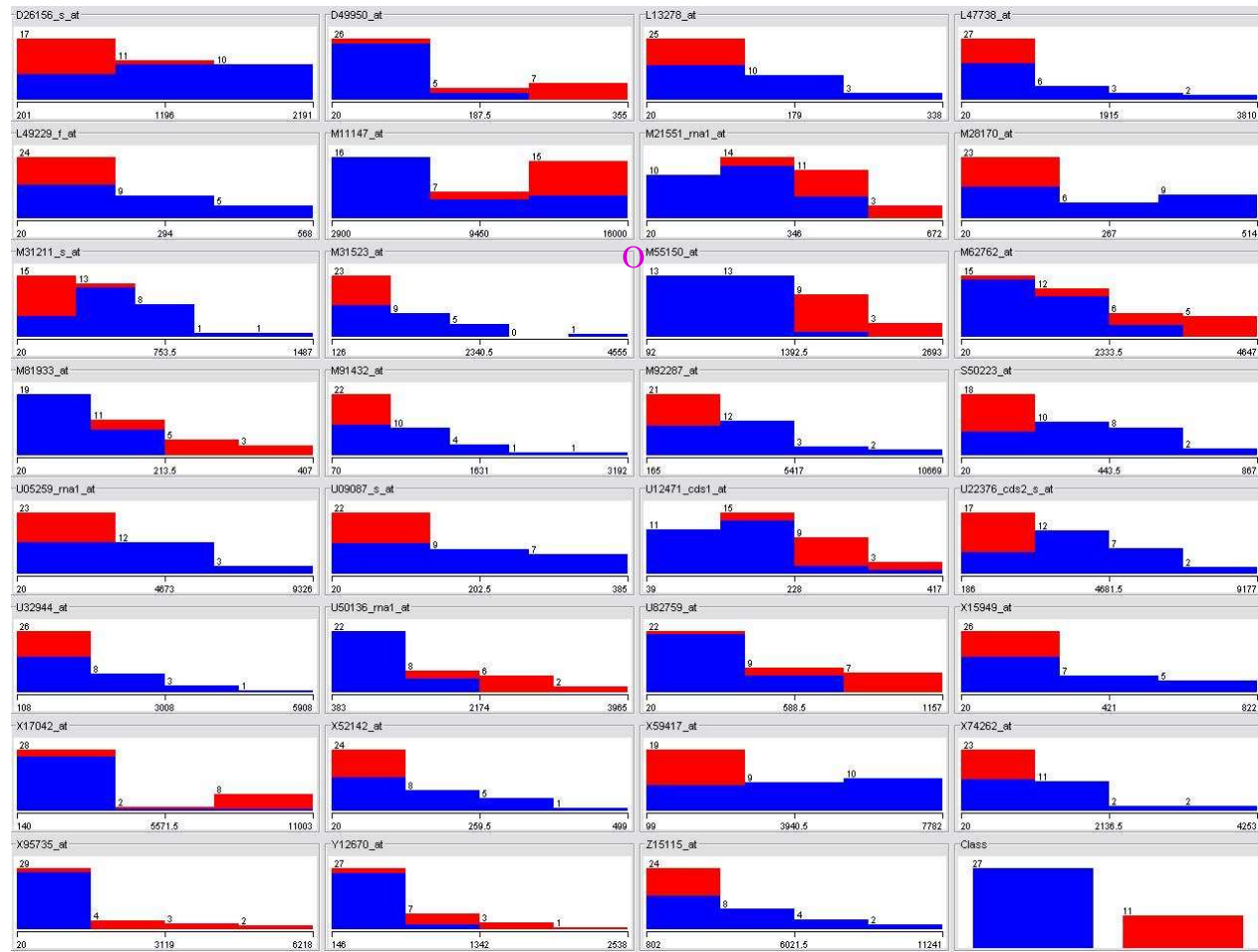
Remarks from Preliminary

It is generally hard to “improve” the mining result from the preliminary analysis.

- Validating the mining result is difficult.
- The training set limits the ability of the study.
- The statistical model may be wrong with the given two types of leukemia.
- Bayes Method is not a good choice here for mining the microarray data.



Attribute Visualization



New Attribute Set

Visualize the 31 attributes, and some attributes are not “helpful” taking Bayes method. Remove them and get a new set of 15 attributes.

D26156_s_at	D49950_at	M11147_at
M21551_rna1_at	M31211_s_at	M55150_at
M62762_at	M81933_at	U12471_cds1_att
U22376_cds2_s_a	U50136_rna1_at	U82759_at
X17042_at	X95735_at	Y12670_at



Mining Result

The results of applying NaïveBayes and BayesNet.

<i>Class</i>	<i>NaïveBaye</i>	<i>BayesNet</i>	<i>Not Agree</i>
ALL	23	26	52, 54, 61 (BN)
AML	11	8	52, 54, 61 (NB)





Discussion

What can the Bayes Method do here?

- Data Preprocessing.
- Attribute Selection.
- Validation with more training samples.





Ending

Questions regarding Analysis results?

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

E-mail: xiao@cs.queens.ca

Thank you

