

Classification rules and descriptive induction

2/12/2010

Voting dataset

Voting dataset

- 435 instances
- 16 attributes
 - 16 nominal attributes
 - 0 numeric attributes
- No target variable
- No missing values

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Open fil... | Open U... | Open D... | Undo | Edit... | Save...

Filter: Choose **None** Apply

Current relation
Relation: voting
Instances: 435 Attributes: 16

Selected attribute
Name: handicapped-infants Type: No...
Missi... 12 (... Distinct: ... Unique: 0 (0...)

Label	Count
n	236
y	187

Class: party (Nom) Visualize All

Attributes: All | None | Invert

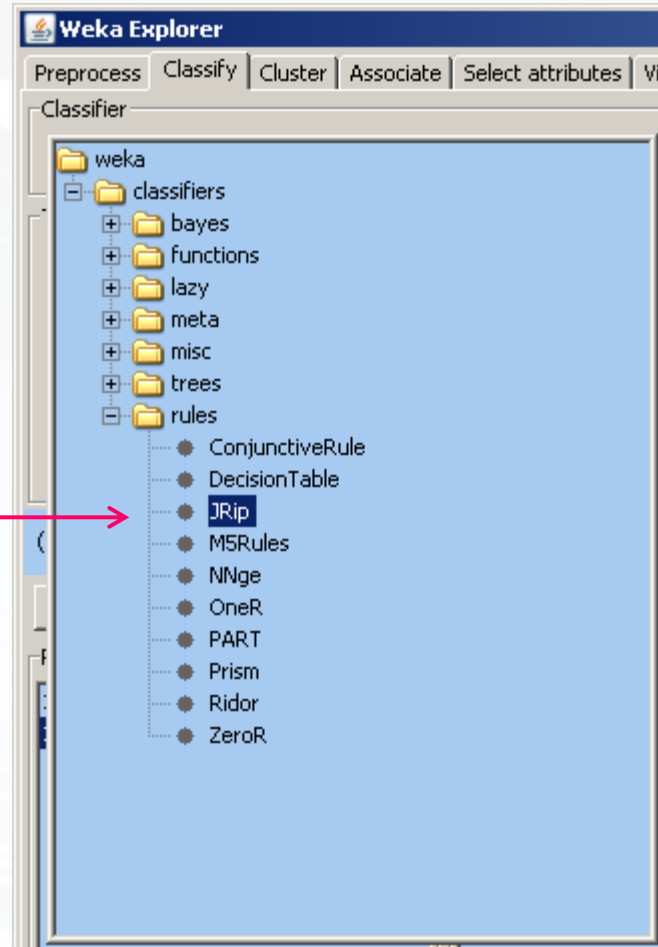
No.	Name
<input checked="" type="checkbox"/>	1 handicapped-infants
<input type="checkbox"/>	2 water-project-cost-sharing
<input type="checkbox"/>	3 adoption-of-the-budget-resolution
<input type="checkbox"/>	4 physician-fee-freeze
<input type="checkbox"/>	5 el-salvador-aid
<input type="checkbox"/>	6 religious-groups-in-schools
<input type="checkbox"/>	7 anti-satellite-test-ban
<input type="checkbox"/>	8 aid-to-nicaraguan-contras
<input type="checkbox"/>	9 mx-missile
<input type="checkbox"/>	10 immigration
<input type="checkbox"/>	11 synfuels-corporation-cutback
<input type="checkbox"/>	12 education-spending
<input type="checkbox"/>	13 superfund-right-to-sue
<input type="checkbox"/>	14 crime
<input type="checkbox"/>	15 duty-free-exports
<input type="checkbox"/>	16 party

Remove

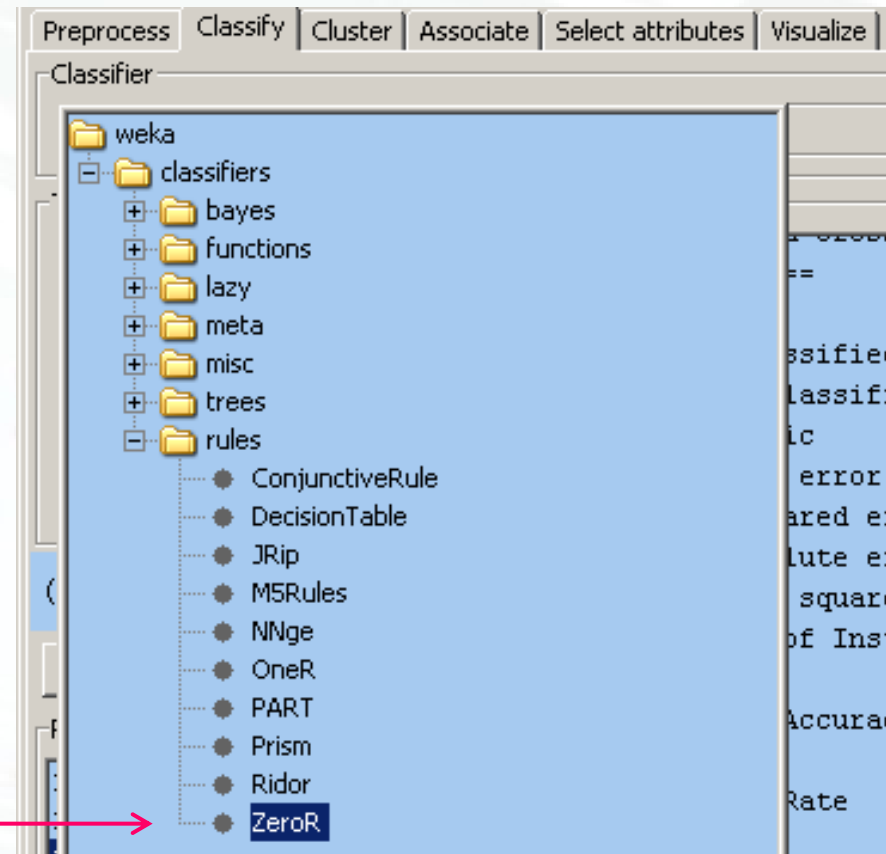
Status: OK Log x 0

Classification rules:

Weka → classifiers → rules → JRip



Baseline classifier: Weka → classifiers → rules → ZeroR



Association rules

Weka → associations → Apriori

①

②

The screenshot shows the Weka Explorer interface with the 'Associate' tab selected. The 'Apriori' algorithm is chosen, and the 'Start' button has been clicked. The 'Associator output' pane displays the following results:

```

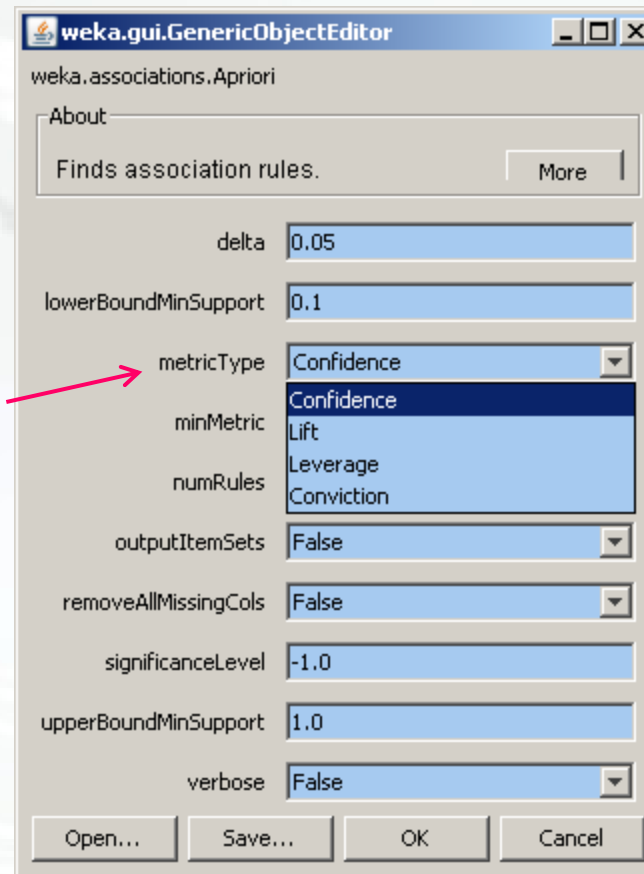
Size of set of large itemsets L(4): 1

Best rules found:

1. adoption-of-the-budget-resolution=y physician-fee-freeze=n 219 ==> party=democrat 219   conf: (1)
2. adoption-of-the-budget-resolution=y physician-fee-freeze=n aid-to-nicaraguan-contras=y 198 ==> party=democrat 198   conf: (1)
3. physician-fee-freeze=n aid-to-nicaraguan-contras=y 211 ==> party=democrat 210   conf: (1)
4. physician-fee-freeze=n education-spending=n 202 ==> party=democrat 201   conf: (1)
5. physician-fee-freeze=n 247 ==> party=democrat 245   conf: (0.99)
6. el-salvador-aid=n party=democrat 200 ==> aid-to-nicaraguan-contras=y 197   conf: (0.99)
7. el-salvador-aid=n 208 ==> aid-to-nicaraguan-contras=y 204   conf: (0.98)
8. adoption-of-the-budget-resolution=y aid-to-nicaraguan-contras=y party=democrat 203 ==> physician-fee-freeze=n 203   conf: (0.97)
9. el-salvador-aid=n aid-to-nicaraguan-contras=y 204 ==> party=democrat 197   conf: (0.97)
10. aid-to-nicaraguan-contras=y party=democrat 218 ==> physician-fee-freeze=n 210   conf: (0.96)
  
```

The 'Result list' on the left shows '16:25:34 - Apriori' as the selected result. The 'Status' bar at the bottom indicates 'OK'.

Quality of association rules



Compare classification and association rules

- Purpose
- Format
- Quality measure
- Ruleset / List of rules
- Exhaustiveness of algorithms

Comparison of classification and association rules

- Purpose
 - Classification rules: classification
 - Association rules: exploratory data analysis, descriptive induction
- Format
 - Both in the format $X \rightarrow Y$
 - Classification rules: Y is a pair “target variable=class”
 - Association rules: both X and Y are itemsets \cong conjunctions of attribute-value pairs
- Quality measure
 - Classification rules: classification accuracy of the ruleset, precision, rule accuracy, weighted relative accuracy
 - Association rules: support, confidence, lift, leverage, conviction
- Ruleset / List of rules
 - Classification rules: can be both: unordered sets of rules or ordered list of rules
 - Association rules: unordered set of rules
- Exhaustiveness of algorithms
 - Classification rules: heuristic algorithms
 - Association rules: exhaustive algorithms, guarantee the optimal results