

Classification rules and descriptive induction

24/11/2009

Voting dataset

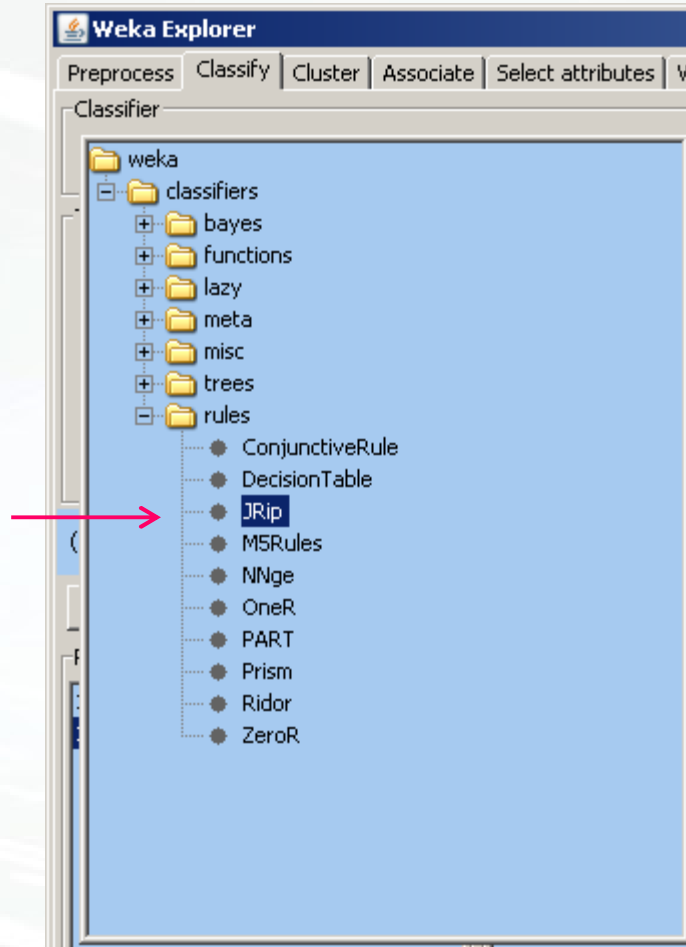
Voting dataset

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

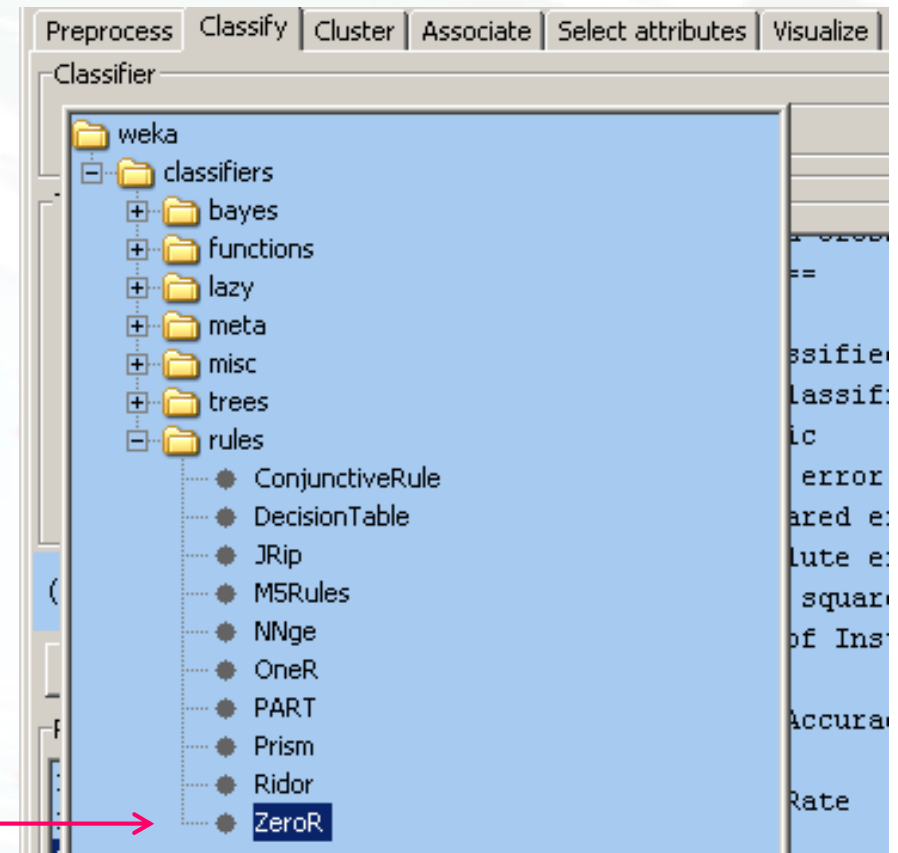
The screenshot shows the Weka Explorer interface for the 'voting' dataset. The 'Attributes' list includes 16 nominal attributes, with 'handicapped-infants' selected. The 'Selected attribute' section shows the distribution of the 'party' class, with 236 instances for 'n' and 187 for 'y'.

Label	Count
n	236
y	187

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 'Associator' dropdown is set to 'Apriori' with parameters: -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0. The 'Start' button is highlighted. The 'Associator output' window displays the following text:

```

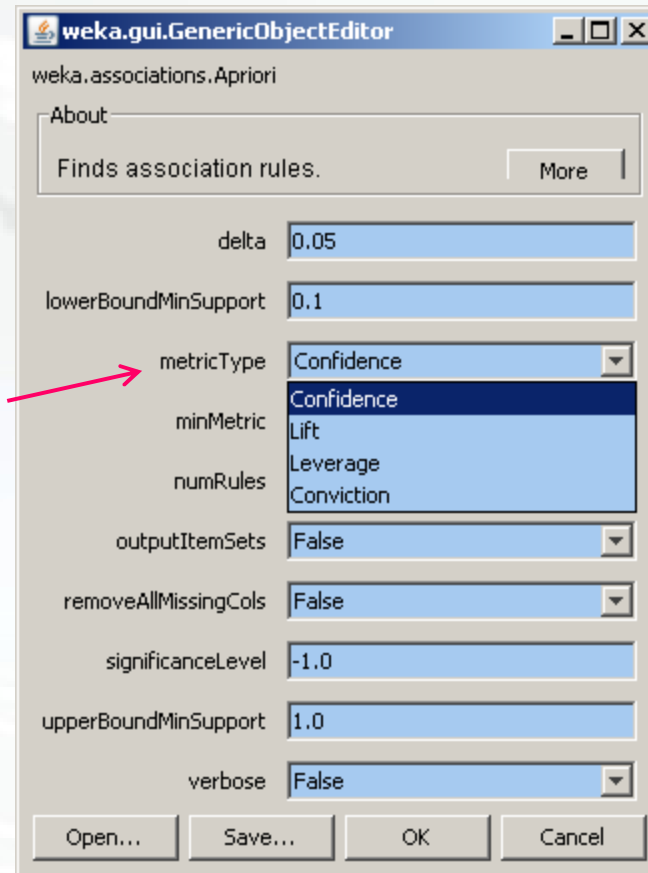
Size of set of large itemsets L(4): 1

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: (0.98)
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.

Quality of association rules



Compare classification and association rules

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

Compare 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