Iterative Optimization of Rule Sets

Jiawei Du
16. November 2010

Prof. Dr. Johannes Fürnkranz
Frederik Janssen
Overview

- REP-Based Algorithms
- RIPPER
- Variants
- Evaluation
- Summary
REP-Based Algorithms

- Split Training Data
- Learn a Rule Set
- Prune the Rule Set
- Learn a Rule
- Prune the Rule
- Check the Rule
- Learn a Rule Set (I-REP*)
  - Split Training Data
  - Learn a Rule
  - Prune the Rule
  - Check the Rule
- Learn a Rule Set (I-REP*)
  - Optimize the Rule Set
    - Get a Rule
    - Generate Variants
    - Choose One Variant
    - Learn Rules (I-REP*)
- Learn a Rule Set (I-REP*)
  - Optimize the Rule Set
    - k times

*k means the number of optimization iterations*
RIPPER
Iterative Optimization of Rule Sets

<table>
<thead>
<tr>
<th>Candidate Rule</th>
<th>Growing Phase</th>
<th>Pruning Phase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Old Rule</td>
<td>Growing a new rule from an empty rule</td>
<td>The pruning heuristic is guided to minimize the error of the single rule</td>
</tr>
<tr>
<td>Replacement</td>
<td>See Old Rule</td>
<td>The pruning heuristic is guided to minimize the error of the entire rule set</td>
</tr>
<tr>
<td>Revision</td>
<td>Further growing the given Old Rule</td>
<td>See Replacement</td>
</tr>
</tbody>
</table>

Selection among the candidate rules based on Minimum Description Length (MDL)

Old Rule
Replacement
Revision

Selection Criterion

Best Rule

Learn a Rule Set (I-REP*)
Optimize the Rule Set

Get a Rule
Generate Variants
Choose One Variant
Learn Rules (I-REP*)

n times

* n means the number of rules in the rule set
1st Variant

New Pruning Method
Candidate Rule Abridgment

Rule: Class = A: C_1, C_2, C_3, C_4

Original Pruning Method
R_1: Class = A: C_1, C_2, C_3 (after 1. Iteration)
R_2: Class = A: C_1, C_2 (after 2. Iteration)
R_3: Class = A: C_1 (after 3. Iteration)

New Pruning Method
R_1’: Class = A: C_2, C_3, C_4
R_2’: Class = A: C_1, C_3, C_4
R_3’: Class = A: C_1, C_2, C_4
R_4’: Class = A: C_1, C_2, C_3 (after 1. Iteration)

* n means the number of rules in the rule set
1st Variant
Search Space
2nd Variant

Simplified Selection Criterion

**Accuracy** instead of **MDL**

\[
\text{MDL} (RS') = DL (RS') - \text{Potentials} (RS')
\]

\[
\text{Potentials} (RS') = \sum \text{Potential}(R'_i) \quad R'_i \in \{RS'\}
\]

\[
\text{Potential}(R'_i) \quad \text{calculates the potential of decreasing the DL of the rule sets if the rule } R'_i \text{ is deleted}
\]

\[
\text{Accuracy}(R_i) = \frac{tp + tn}{P + N} \quad R_i \in \{\text{OldRule, Replacement, Revision}\}
\]

- **tp** means the number of positive examples covered by the relevant rule
- **tn** means the number of negative examples that are not covered by the relevant rule
- **P** and **N** mean the total number of positive and negative examples in the training set

\[\sum \quad n \text{ times}\]

\[\ast n \text{ means the number of rules in the rule set}\]
Evaluation

- **Data Sets**
  
  20 real data sets selected from the UCI repository
  
  - 9 data sets  (type categorical)
  - 4 data sets  (type numerical)
  - 7 data sets  (type mixed)

- **Evaluation Method**

  10-fold stratified cross-validation
  
  - run 10 times on each data set
  
  - training set  90%
  - testing set    10%
Evaluation

RIPPER (SeCoRIP)

- The correctness of rule sets is increased *(the percentage of the correctly classified examples in the testing set)*
- The size of rule set is decreased
- The number of conditions in each rule is decreased

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>AvgCorr.</th>
<th>Profit</th>
</tr>
</thead>
<tbody>
<tr>
<td>SeCoRIP_0</td>
<td>86.19</td>
<td>-</td>
</tr>
<tr>
<td>SeCoRIP_1</td>
<td>87.56</td>
<td>1.59%</td>
</tr>
<tr>
<td>SeCoRIP_2</td>
<td>87.61</td>
<td>0.06%</td>
</tr>
<tr>
<td>SeCoRIP_3</td>
<td>87.53</td>
<td>-0.08%</td>
</tr>
<tr>
<td>SeCoRIP_4</td>
<td>87.64</td>
<td>0.12%</td>
</tr>
<tr>
<td>SeCoRIP_5</td>
<td>87.45</td>
<td>-0.21%</td>
</tr>
</tbody>
</table>

**Profit**

\[
\text{Profit}_{i+1} = \frac{\text{AvgCorr}_{i+1} - \text{AvgCorr}_i}{\text{AvgCorr}_i} \quad i \in \{0, 1, 2, 3, 4\}
\]
Evaluation

RIPPER (Convergence of SeCoRIP)

Group A
- The maximal value appears at the x-axis
- Optimizations $= 0$
- These points converge to a definite point
- The relevant data sets contain only nominal attributes

Group B
- The maximal value mainly appears at the x-axis
- Optimizations $\in \{1, 2\}$
- These points converge to a definite point
- The relevant data sets contain more nominal attributes than numeric ones
Evaluation

**RIPPER (Convergence of SeCoRIP)**

- The maximal value mainly appears at the x-axis $\text{Optimizations} \in \{5, 6, 7\}$
- These points converge to a definite point

**Group C**

**Group D**

- The points of the lines show a upward trend at the x-axis $\text{Optimizations} \in \{8, 9, 10\}$
- The signal of convergence is not observable
- The relevant data sets contain more numeric attributes than nominal ones
Evaluation

RIPPER (Convergence of SeCoRIP)

- N (nominal attributes) > N (numerical attributes)
  - the accuracy of the optimized rule sets often converge to a definite value with the increasing of the number of optimization iterations
  - the definite value here is usually not the maximum or minimum value obtained so far
- N (nominal attributes) < N (numerical attributes)
  - The value of the correctness keeps an upward trend with the increasing of the number of optimization iterations
  - The signal of convergence cannot be obviously detected
Evaluation

RIPPER (SeCoRIP)

- The correctness of rule sets is increased
- The size of rule set is decreased (the sum of all rules in the constructed rule sets)
- The number of conditions in each rule is decreased (the sum of all conditions / the size of rule set)

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>AvgRules</th>
<th>AvgCond. in one Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>SeCoRIP_0</td>
<td>8.75</td>
<td>1.94</td>
</tr>
<tr>
<td>SeCoRIP_1</td>
<td>7.35</td>
<td>1.65</td>
</tr>
<tr>
<td>SeCoRIP_2</td>
<td>7.25</td>
<td>1.69</td>
</tr>
<tr>
<td>SeCoRIP_3</td>
<td>7.40</td>
<td>1.73</td>
</tr>
<tr>
<td>SeCoRIP_4</td>
<td>7.55</td>
<td>1.73</td>
</tr>
<tr>
<td>SeCoRIP_5</td>
<td>7.50</td>
<td>1.73</td>
</tr>
</tbody>
</table>
Evaluation

1st Variant (SeCoRIP*)

- The new pruning method will have no obvious effect on the rule sets whose rules contain too few conditions
- Sometimes the constructed Abridgement is the same as the candidate rule Revision or even the original Old Rule

![Diagram](image)

- The correctness of the rule sets can be well improved when the relevant rules normally contain more than three conditions
Evaluation

2nd Variant (SeCoRIP')
Evaluation

2\textsuperscript{nd} Variant (SeCoRIP’)

Compare to SeCoRIP:
- The correctness of the constructed rule sets are often worse
- The difference can be reduced with the increasing of the number of optimization iterations
- Several data sets cannot be well processed
- The number of rules and conditions can also be decreased

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>AvgRules</th>
<th>AvgCond. in one Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>SeCoRIP_0</td>
<td>8.75</td>
<td>1.94</td>
</tr>
<tr>
<td>SeCoRIP_1</td>
<td>7.05</td>
<td>1.70</td>
</tr>
<tr>
<td>SeCoRIP_2</td>
<td>7.00</td>
<td>1.72</td>
</tr>
<tr>
<td>SeCoRIP_3</td>
<td>7.25</td>
<td>1.74</td>
</tr>
<tr>
<td>SeCoRIP_4</td>
<td>7.05</td>
<td>1.74</td>
</tr>
<tr>
<td>SeCoRIP_5</td>
<td>7.25</td>
<td>1.77</td>
</tr>
</tbody>
</table>
Summary

- **RIPPER** *(postprocessing phase)*
  - The correctness of rule sets is increased
  - The results often converge to a definite value
  - Better handling of the data sets which contain more numeric attributes
  - The number of rules and conditions is decreased

- **1st Variant** *(new pruning method)*
  - Not suitable for the rule sets whose rules contain too few conditions
  - Taking positive effect on the rule sets whose rules contain sufficient number of conditions

- **2nd Variant** *(simplified selection criterion)*
  - Remaining the features of the original version
  - The results are not as good as the original version
  - The original selection criterion MDL is not easily replaceable
Thank you for your attention!