Iterative Optimization of Rule Sets

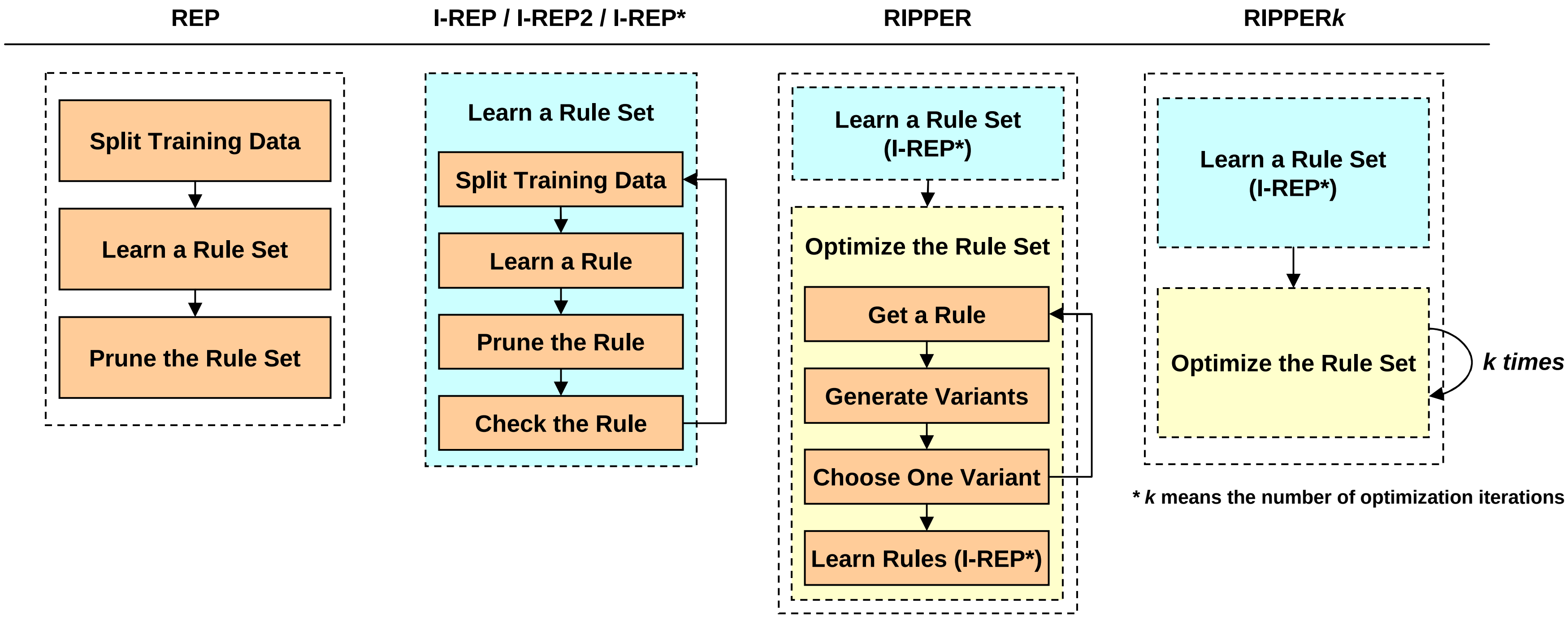
Jiawei Du
16. November 2010

Prof. Dr. Johannes Fürnkranz
Frederik Janssen

Overview

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

REP-Based Algorithms

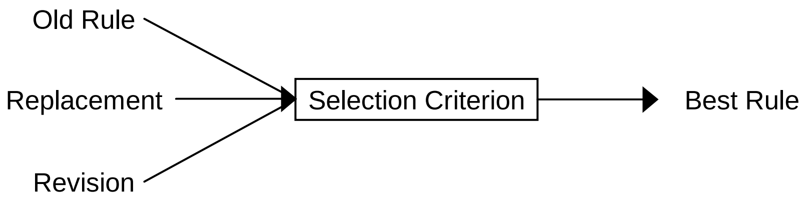


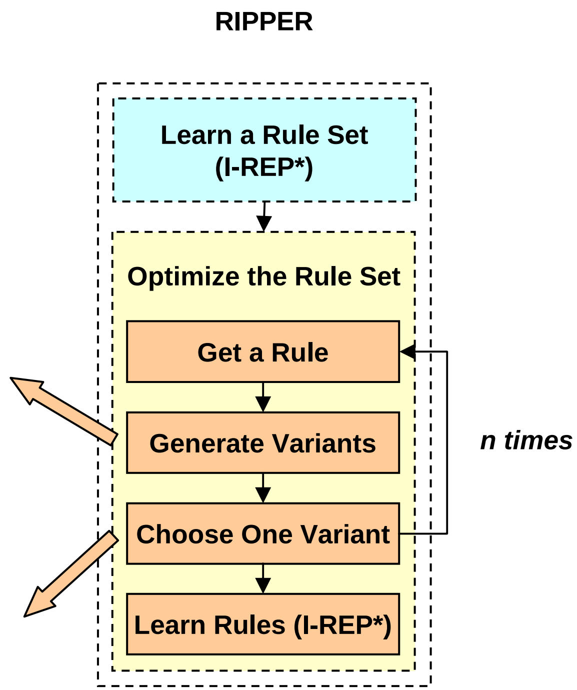
RIPPER

Iterative Optimization of Rule Sets

Candidate Rule	Growing Phase	Pruning Phase
Old Rule	Growing a new rule from an empty rule	The pruning heuristic is guided to minimize the error of the single rule
Replacement	See Old Rule	The pruning heuristic is guided to minimize the error of the entire rule set
Revision	Further growing the given Old Rule	See Replacement

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





*** *n* means the number of rules in the rule set**

1st Variant

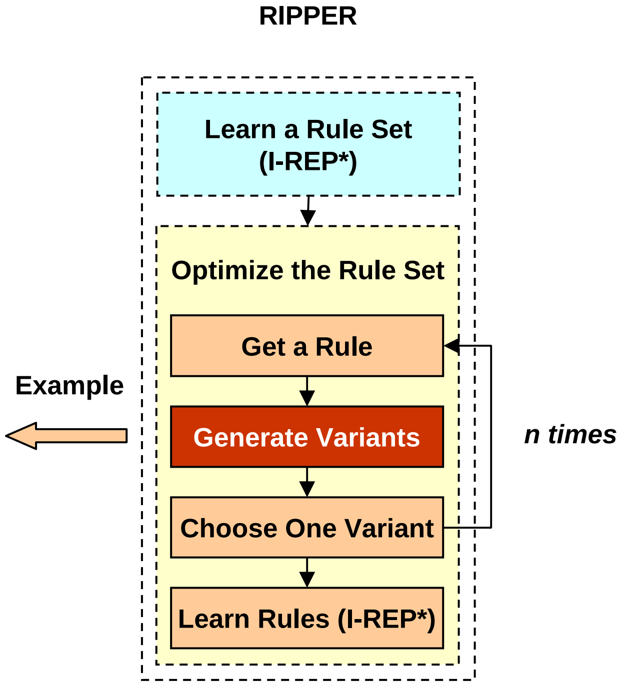
New Pruning Method

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)

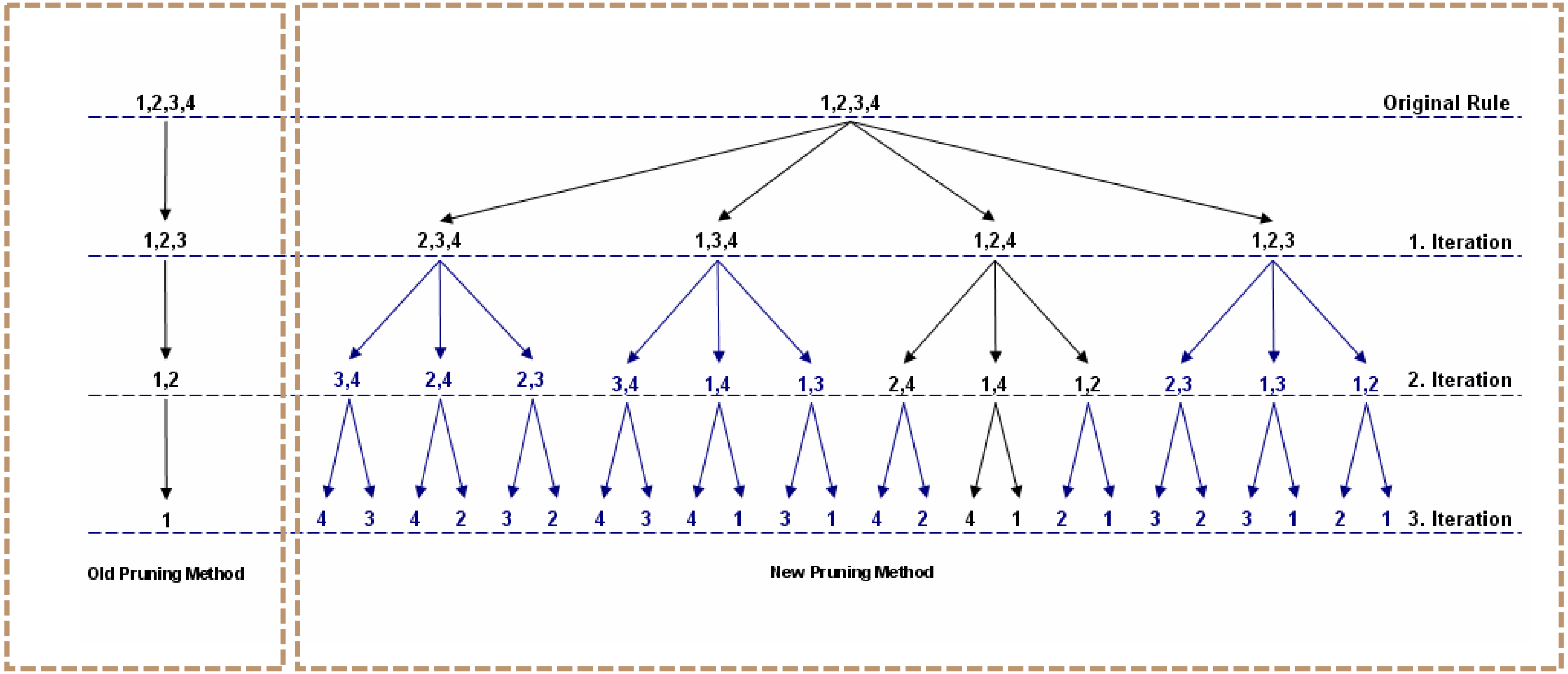
New Pruning Method
R_1': Class = A: C_2, C_3, C_4
R_2' Class = A: C_1, C_3, C_4
R_3': Class = A: C_1, C_2, C_4
R_4': Class = A: C_1, C_2, C_3 (after 1. Iteration)



* *n* means the number of rules in the rule set

1st Variant

Search Space



2nd Variant

Simplified Selection Criterion

Accuracy instead of **MDL**

MDL (RS') = DL (RS') – Potentials (RS')

Potentials (RS') $= \sum Potential(R_i') \quad R_i' \in \{RS'\}$

$Potential\,(R_i')$ calculates the potential of decreasing the *DL* of the rule sets if the rule R_i' is deleted

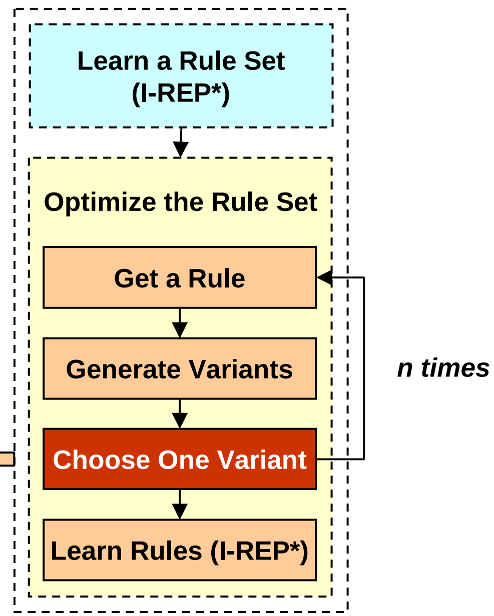
$$Accuracy(R_i) = \frac{tp + tn}{P + N} \qquad R_i \in \{\text{OldRule}, \text{Replacement}, \text{Revision}\}$$

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

RIPPER



*** *n* 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)

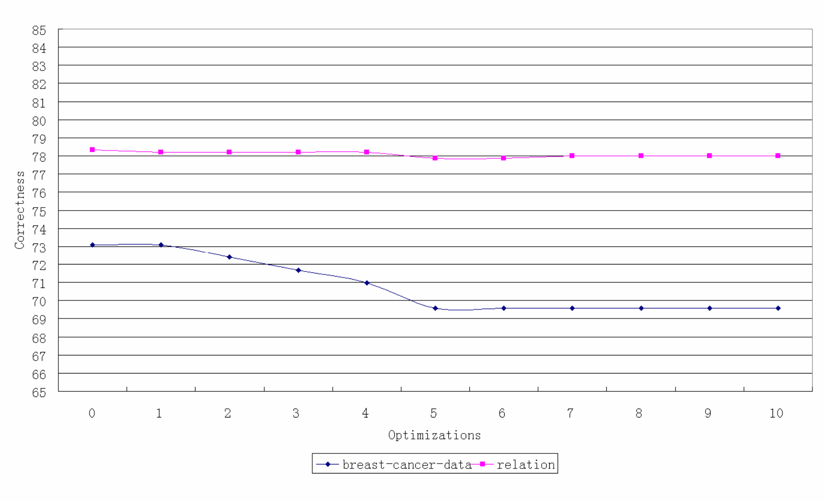
- The correctness of rule sets is increased *(the percentage of the correctly classified examples in the testing set)*
- The size of rule set is decreased
- The number of conditions in each rule is decreased

Algorithm	AvgCorr.	Profit
SeCoRIP_0	86.19	-
SeCoRIP_1	87.56	1.59%
SeCoRIP_2	87.61	0.06%
SeCoRIP_3	87.53	-0.08%
SeCoRIP_4	87.64	0.12%
SeCoRIP_5	87.45	-0.21%

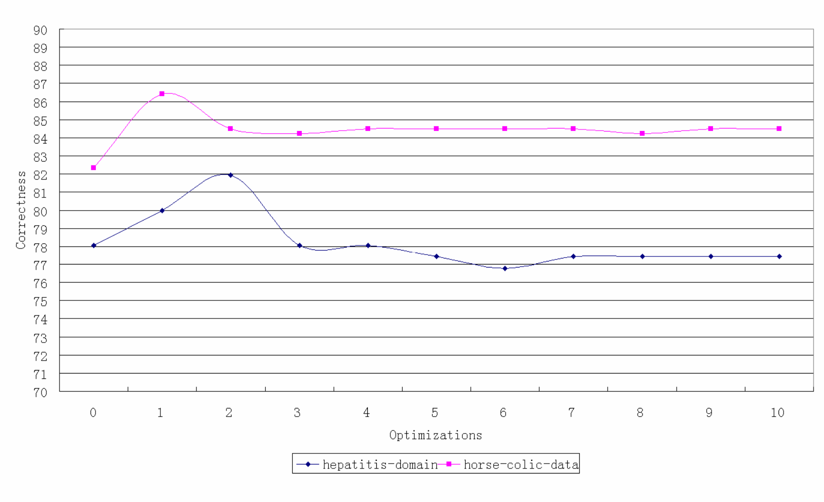
$$\text{Profit}_{(i+1)} = \frac{\text{AvgCorr}_{(i+1)} - \text{AvgCorr}_i}{\text{AvgCorr}_i} \qquad i \in \{0,1,2,3,4\}$$

Evaluation

RIPPER (Convergence of SeCoRIP)



Group A



Group B

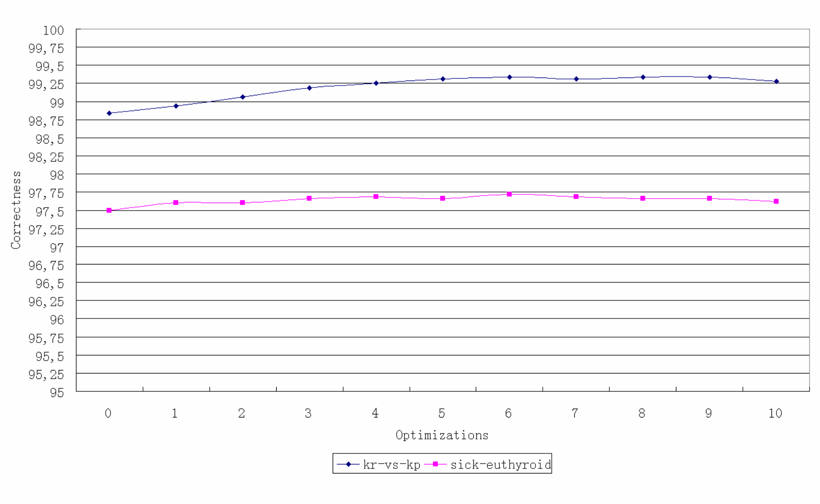
- The maximal value appears at the x-axis *Optimizations* $= 0$
- These points converge to a definite point
- The relevant data sets contain only nominal attributes

- 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

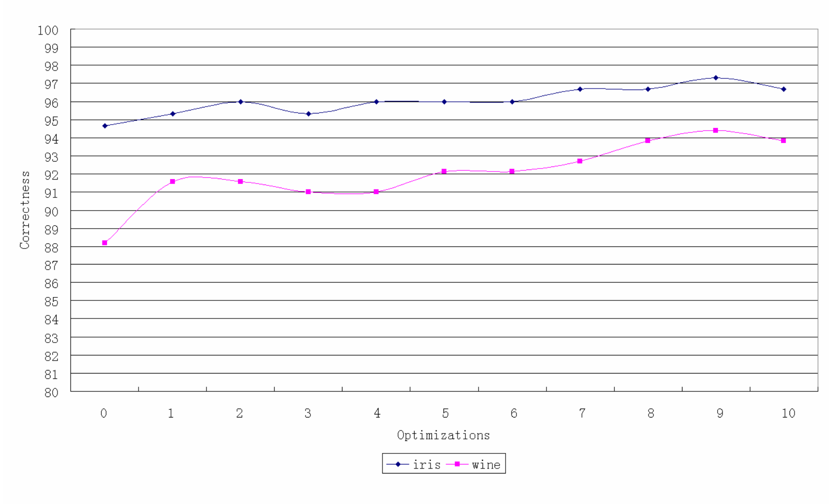
TECHNISCHE UNIVERSITÄT DARMSTADT

Evaluation

RIPPER (Convergence of SeCoRIP)



Group C



Group D

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

- The points of the lines show a upward trend at the x-axis *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)*

Algorithm	AvgRules.	AvgCond. in one Rule
SeCoRIP_0	8.75	1.94
SeCoRIP_1	7.35	1.65
SeCoRIP_2	7.25	1.69
SeCoRIP_3	7.40	1.73
SeCoRIP_4	7.55	1.73
SeCoRIP_5	7.50	1.73

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**

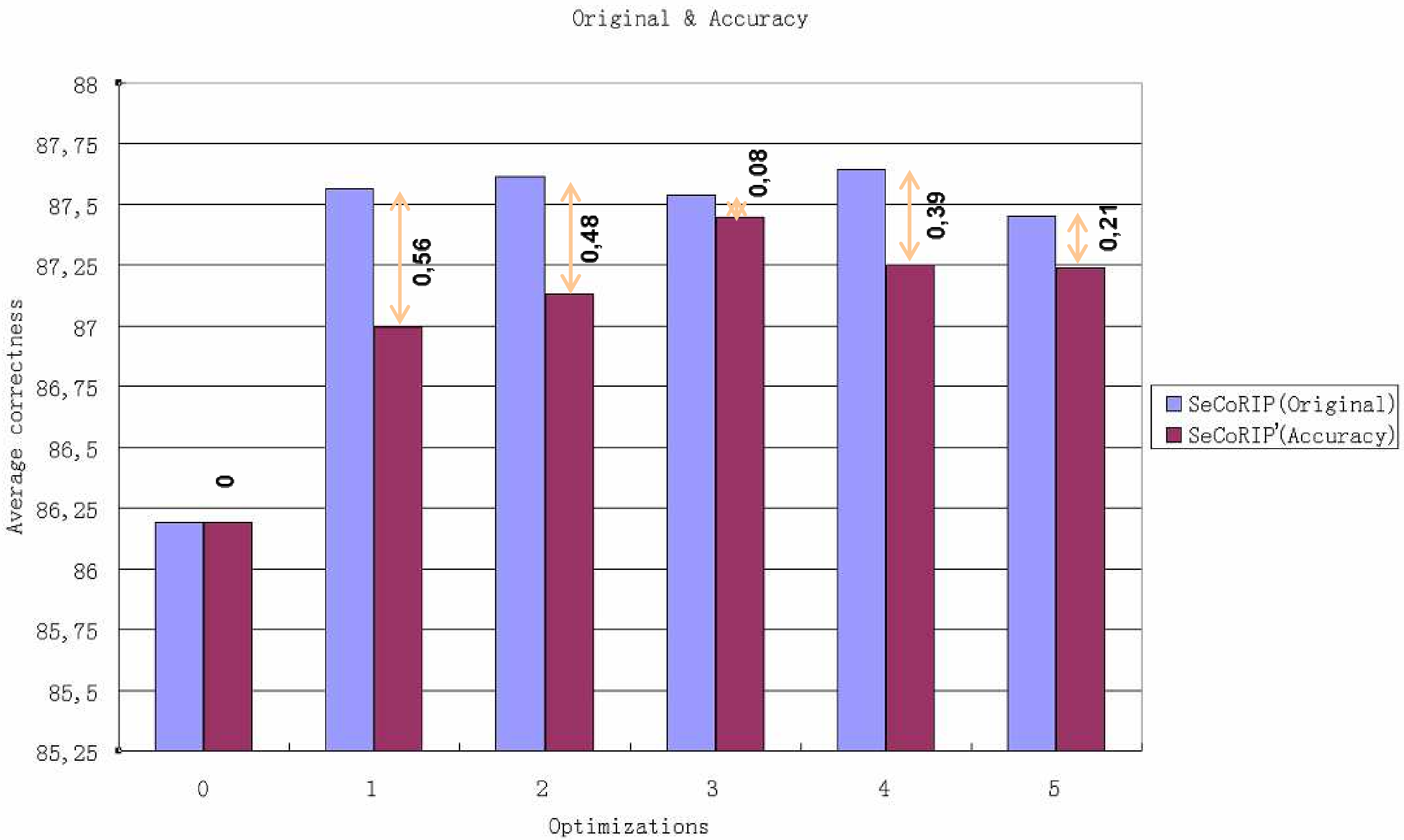
R: Class = A: C_1, C_2 → R': Class = A: C_1

R: Class = A: C_1, C_2 → R': Class = A: C_1, C_2

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

Evaluation

2nd Variant (SeCoRIP')



TECHNISCHE UNIVERSITÄT DARMSTADT

Evaluation

2nd 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

Algorithm	AvgRules.	AvgCond. in one Rule
SeCoRIP_0’	8.75	1.94
SeCoRIP_1’	7.05	1.70
SeCoRIP_2’	7.00	1.72
SeCoRIP_3’	7.25	1.74
SeCoRIP_4’	7.05	1.74
SeCoRIP_5’	7.25	1.77

Summary

- RIPPER *(postprocessing phase)*
  - The correctness of rule sets is increased
  - The results often converge to a definite value
  - Better handling 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!

TECHNISCHE UNIVERSITÄT DARMSTADT