RULES-6: a simple rule induction algorithm for handling large data sets

D. T. Pham and A. A. Afify

Seminarvortrag: Jan Meub

TECHNISCHE UNIVERSITÄT DARMSTADT

RULe Extraction System

Version 6

RULES algorithm family

TECHNISCHE UNIVERSITÄT DARMSTADT

Deveoped over 20 years by D.T. Pham et al.

1:	Pham and Aksoy	(93)
2:	Pham and Aksoy	(95)
3:	Pham and Aksoy	(95)
3+	Pham and Dimov	(97)
4:	Pham and Dimov	(97)
5:	Pham, Bigot and Dimov	(03)
F:	Pham, Bigot and Dimov	(06)
6:	Pham and Afify	(05)
7:	Pham and Shehzad	(10)
8:	Pham and Pham	(12)

RULES-3 Plus - Problems

TECHNISCHE UNIVERSITÄT DARMSTADT

Complete and consistent on training data

- Overfitting
- Noise

H measure is complex and not accurate enough

Equal-width discretisation is inefficient

RULES-6

TECHNISCHE UNIVERSITÄT DARMSTADT

Inductive seperate and conquer rule set learning

```
RuleSet = ∅
While any example in TrainingSet is not covered
    s = any uncovered example (seed)
    InduceRule(s, TrainingSet, BeamWidth)
    Mark examples covered by Rule as covered
    RuleSet = RuleSet ∪ {Rule}
Return RuleSet
```

InduceOneRule(s, TrainingSet, BeamWidth: w)

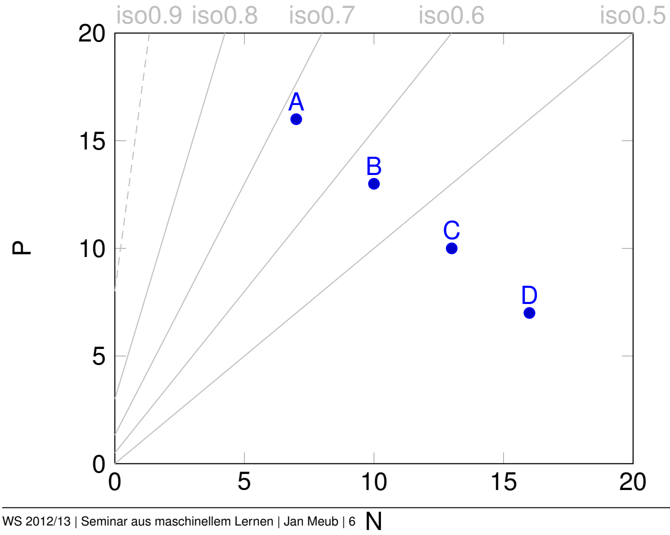
TECHNISCHE UNIVERSITÄT DARMSTADT

pruned general to specific beam search

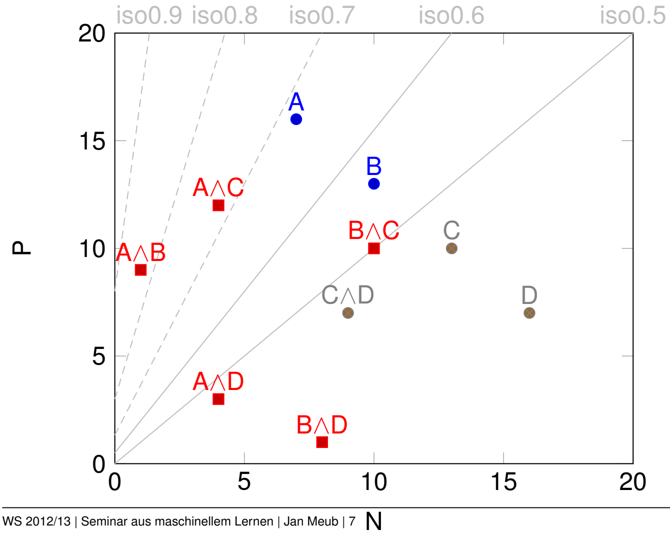
PartialRules = NewPartialRules = $\varnothing$
BestRule = most general rule (no conditions)
PartialRules = PartialRules $\cup$ {BestRule}
While (PartialRules $\neq \varnothing$)
    $\forall Rule \in PartialRules$:
        Specialise(Rule, s) $\Rightarrow$ NewPartialRules, BestRule
    Prune rules that cannot improve from NewPartialRules
    $\forall Rule \in NewPartialRules$:
        Rule.ValidValues -= Parent(Rule).InvalidValues
    PartialRules = w best Rules from NewPartialRules (no duplicates)
Return BestRule

BeamSearch

TECHNISCHE UNIVERSITÄT DARMSTADT

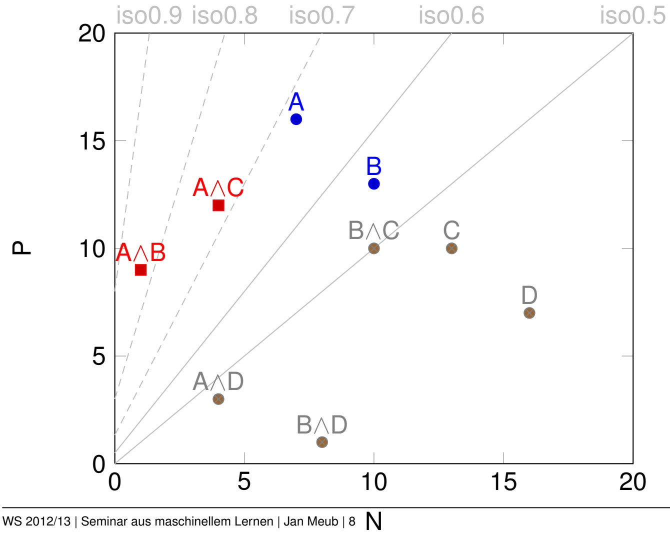


BeamSearch



iso0.9
iso0.8
iso0.7
iso0.6
iso0.5
A
B
A∧C
B∧C
C
A∧B
C∧D
D
A∧D
B∧D
P
N
0
5
10
15
20
0
5
10
15
20

BeamSearch



Specialise (Rule, s)

TECHNISCHE UNIVERSITÄT DARMSTADT

For each nominal attribute A_i that does not appear in Rule
  If $v_{is} \in$ Rule.ValidValues (v_{is} value of A_i in s) **Then**
    NewRule = Rule $\wedge$ (A_i = v_{is})
    If NewRule.Score > BestRule.Score **Then**
      BestRule = NewRule
    If stopSpecialisation
      Then Parent(NewRule).InvalidValues += v_{is}
      Else NewPartialRules += NewRule

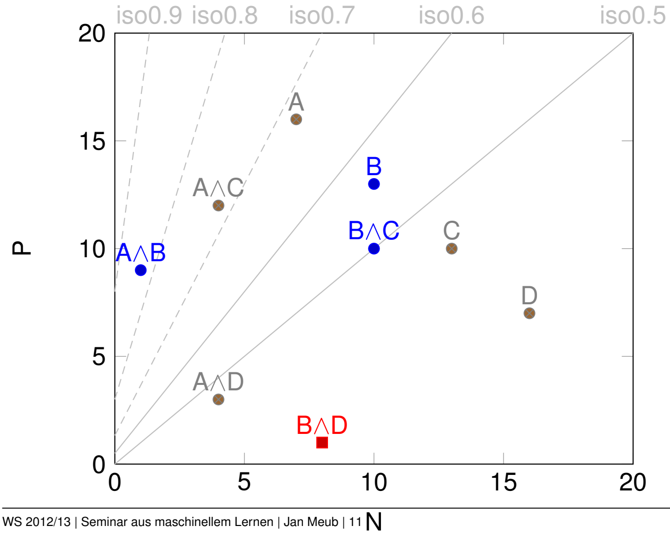
stopSpecialisation
Pruning conditions 1-3

CoveredPositives(NewRule) $\leq$ MinPositives
Or
CoveredNegatives(Rule) - CoveredNegatives(NewRule) $\leq$ MinNegatives
Or
Consistency(NewRule) = 100%

MinPositives and MinNegatives:
- Empirical evidence suggest values between 1 and 5
- Smaller values if noise is low

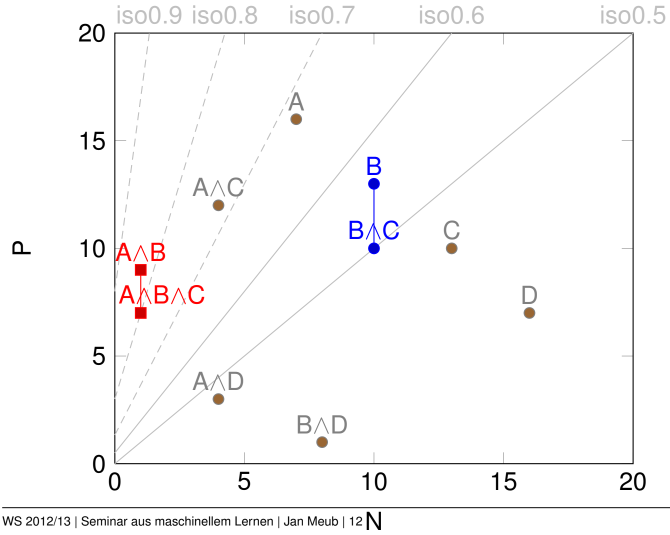
CovP(NewRule) $\leq$ MinP
Pruning 1

TECHNISCHE UNIVERSITÄT DARMSTADT



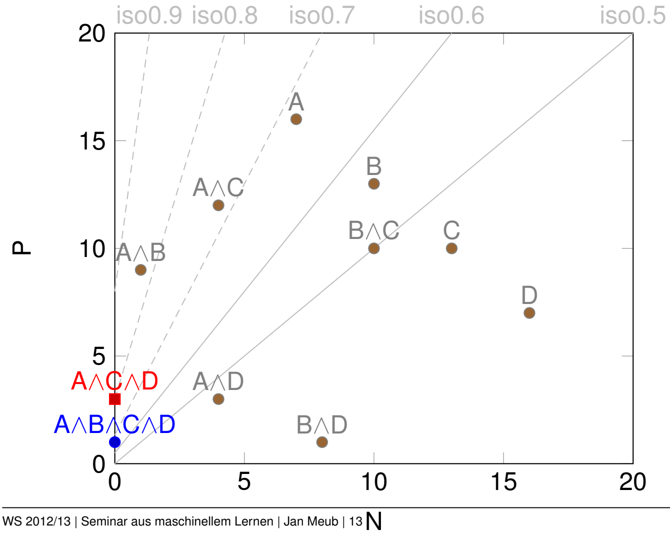
CovN(Rule) - CovN(NewRule) $\leq$ MinN
Pruning 2

TECHNISCHE UNIVERSITÄT DARMSTADT



Consistency(NewRule) = 100%
Pruning 3

TECHNISCHE UNIVERSITÄT DARMSTADT



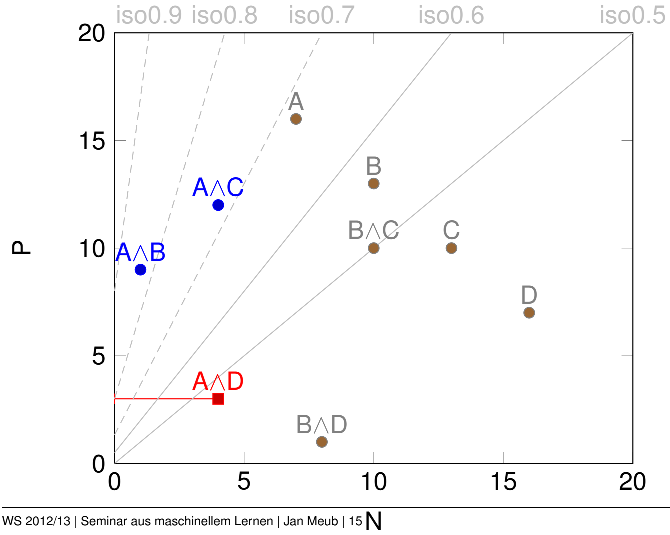
Prune rules that cannot improve
Pruning 4

For each Rule $\in$ NewPartialRules
  If Rule.OptimisticScore < BestRule.Score **Then**
    NewPartialRules -= Rule
    Parent(NewRule).InvalidValues += v_{is}

OptimisticScore: Assume CoveredNegatives = 0
$\Rightarrow$ max score any specialisation of Rule can get

Prune rules that cannot improve
Pruning 4

TECHNISCHE
UNIVERSITÄT
DARMSTADT



Rule.Score
Metric: m-probability-estimate

Rule.Score = $\frac{n_{class}+mP_0(C_t)}{n_{covered}+m}$

n_{class} = | positive examples covered |
$n_{covered}$ = | total examples covered |
$P_0(C_t)$ = a priori probability of class C_t
m = domain dependant parameter, increase with noise

For m = k = |classes| and P_0 = uniform = $\frac{1}{k}$
$\rightarrow$ Laplace estimate ($\frac{n_{class}+1}{n_{covered}+k}$)

In RULES-6:
m = k, $P_0(C_t) = \frac{N_{C_t}}{N}$

Evaluation of the search-space pruning rules

Evaluation on 40 data sets from the University of California at Irvine (UCI) repository of machine learning databases.

Total reduction of search space: 80% (up to 99.4%)

Total increase in accuracy: 8.3%

All four pruning methods can decrease accuracy.

MinNegatives seems to have the highest impact

Comparison with RULES-3 Plus

- Number of Rules: -88.5%
- Number of Conditions: -95.7%
- Number of Evaluations: -94.8%
- Execution Time: -97.6%

- Slight increase in accuracy (inc: 25 | dec:11)

Comparison with C5.0

TECHNISCHE UNIVERSITÄT DARMSTADT

- Almost identical performance in accuracy and number of rules
- Fewer rules in 30 datasets for C5.0

Evaluation of different descretisation methods

TECHNISCHE UNIVERSITÄT DARMSTADT

Afify, A. A. Design and analysis of scalable rule induction systems.
PhD Thesis, University of Wales Cardiff, School of Engineering,
Systems Engineering Division, Cardiff, UK, 2004.

“showed that the performance of the RULES-6 algorithm significantly improved when continuous valued attributes were discretized using”

Fayyad, U. M. and Irani, K. B. Multi-interval discretization of continuous-valued attributes for classification.
In Proceedings of the 13th International Joint Conference on Artificial intelligence, Chambery, France, 1993, pp. 1022-1027.

Future Work

TECHNISCHE UNIVERSITÄT DARMSTADT

- Additional prepruning
- Use of postpruning
- Integrated discretization

Future Work
RULES-7

Khurram Shehzad
New Rule Induction Algorithm with improved Noise Tolerance and Scalability.
Ph.D. thesis, University of Wales Cardiff. 2010

Khurram Shehzad
EDISC: A Class-Tailored Discretization Technique for Rule-Based Classification
IEEE Transactions on Knowledge and Data Engineering,
vol. 24, no. 8, pp. 1435-1447. 2012

Khurram Shehzad
Simple Hybrid and Incremental Postpruning Techniques for Rule Induction
IEEE Transactions on Knowledge and Data Engineering,
vol. 25, no. 2, pp. 476-480, Feb. 2013

Future Work
RULES-8

Dinh Trung Pham
A Novel Rule Induction Algorithm with Improved Handling of Continuous Valued Attributes
Ph.D. thesis, University of Wales Cardiff. 2012

Sources

D.T. Pham and A.A. Afify, 2005
RULES-6: A Simple Rule Induction Algorithm for Supporting Decision Making
Proc. 31st Ann. Conf. IEEE Industrial Electronics Soc. (IECON), pp. 2184-2189

Questions?

?