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

D. T. Pham and A. A. Afify
Seminarvortrag: Jan Meub

RULe Extraction System
Version 6
RULES algorithm family

Developed 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

Complete and consistent on training data
  ▶ Overfitting
  ▶ Noise

H measure is complex and not accurate enough

Equal-width discretisation is inefficient
Inductive separate and conquer rule set learning

\[
\text{RuleSet} = \emptyset \\
\textbf{While} \text{ any example in TrainingSet is not covered} \\
\quad s = \text{any uncovered example (seed)} \\
\quad \text{InduceRule}(s, \text{TrainingSet}, \text{BeamWidth}) \\
\quad \text{Mark examples covered by Rule as covered} \\
\quad \text{RuleSet} = \text{RuleSet} \cup \{\text{Rule}\} \\
\textbf{Return} \text{ RuleSet}
\]
InduceOneRule(s, TrainingSet, BeamWidth: w)

pruned general to specific beam search

\[
\begin{align*}
\text{PartialRules} &= \text{NewPartialRules} = \emptyset \\
\text{BestRule} &= \text{most general rule (no conditions)} \\
\text{PartialRules} &= \text{PartialRules} \cup \{\text{BestRule}\}
\end{align*}
\]

\textbf{While} (\text{PartialRules} \neq \emptyset)

\[
\forall Rule \in \text{PartialRules}:
\]

- Specialise\(\text{(Rule, s)}\) \Rightarrow \text{NewPartialRules, BestRule}

- Prune rules that cannot improve from \text{NewPartialRules}

\[
\forall Rule \in \text{NewPartialRules}:
\]

- Rule.ValidValues -= Parent\(\text{(Rule)}.InvalidValues\)

\text{PartialRules} = w \text{ best Rules from } \text{NewPartialRules (no duplicates)}

\textbf{Return} \text{BestRule}
BeamSearch

A
B
C
D
BeamSearch

A \land B
A \land C
B \land C
A \land D
B \land D
C \land D

iso0.5 iso0.6 iso0.7 iso0.8 iso0.9
BeamSearch

The diagram illustrates the relationships between different elements labeled A, B, C, and D. Points correspond to combinations of these elements. For example, A ∧ B, A ∧ C, A ∧ D, B ∧ C, C, and D are shown on the graph. The graph uses lines labeled iso0.5, iso0.6, iso0.7, iso0.8, and iso0.9 to represent different iso-contours of a metric. Points A and B are marked with blue circles, while points A ∧ B, A ∧ C, and A ∧ D are marked with red squares. Points B ∧ C, C, and D are marked with brown circles.
Specialise (Rule, s)

For each nominal attribute $A_i$ that does not appear in Rule

If $v_{is} \in \text{Rule.ValidValues}$ ($v_{is}$ value of $A_i$ in s) Then

NewRule = Rule $\land$ ($A_i = v_{is}$)

If NewRule.Score > BestRule.Score Then

BestRule = NewRule

If stopSpecialisation

Then Parent(NewRule).InvalidValues += $v_{is}$

Else NewPartialRules += NewRule
stopSpecialisation
Pruning conditions 1-3

\[
\text{CoveredPositives}(\text{NewRule}) \leq \text{MinPositives} \\
\text{Or} \\
\text{CoveredNegatives}(\text{Rule}) - \text{CoveredNegatives}(\text{NewRule}) \leq \text{MinNegatives} \\
\text{Or} \\
\text{Consistency}(\text{NewRule}) = 100\%
\]

\text{MinPositives and MinNegatives:}
 Empirical evidence suggest values between 1 and 5
 Smaller values if noise is low
CovP(NewRule) \leq \text{MinP}

Pruning 1
CovN(Rule) - CovN(NewRule) ≤ MinN

Pruning 2
Consistency(NewRule) = 100%
Pruning 3

\[ A \land B \land C \land D \]
\[ A \land C \land D \]
\[ A \land B \land C \land D \]
\[ A \land D \]
\[ B \land C \land D \]
\[ B \land D \]

iso0.5 iso0.6 iso0.7 iso0.8 iso0.9
Prune rules that cannot improve

Pruning 4

For each Rule ∈ NewPartialRules
   If Rule.OptimisticScore < BestRule.Score Then
      NewPartialRules -= Rule
      Parent(NewRule).InvalidValues += v_i

OptimisticScore: Assume CoveredNegatives = 0
   ⇒ max score any specialisation of Rule can get
Prune rules that cannot improve
Pruning 4
Rule.Score
Metric: m-probability-estimate

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

n_{\text{class}} = |\text{positive examples covered}|

n_{\text{covered}} = |\text{total examples covered}|

P_0(C_t) = \text{a priori probability of class } C_t

m = \text{domain dependant parameter, increase with noise}

For m = k = |\text{classes}| and P_0 = \text{uniform} = \frac{1}{k}

→ \text{Laplace estimate} (\frac{n_{\text{class}} + 1}{n_{\text{covered}} + k})

In RULES-6:

m = k, \quad 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

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


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

Future Work

▶ Additional prepruning
▶ Use of postpruning
▶ Integrated discretization
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
D.T. Pham and A.A. Afify, 2005
RULES-6: A Simple Rule Induction Algorithm for Supporting Decision Making
Questions?