Evaluating a Rule Evaluation Support Method with Learning Models Based on Objective Rule Evaluation Indices

- A Case Study with a Meningitis Data Mining Results-

Hidenao Abe[1), S. Tsumoto[1), M. Ohsaki[2), T. Yamaguchi[3)

Dept. of Medical Informatics, Shimane University, School of Medicine, Japan[1)

Faculty of Engineering, Doshisha University, Japan[2)

Faculty of Science and Technology, Keio University, Japan[3)

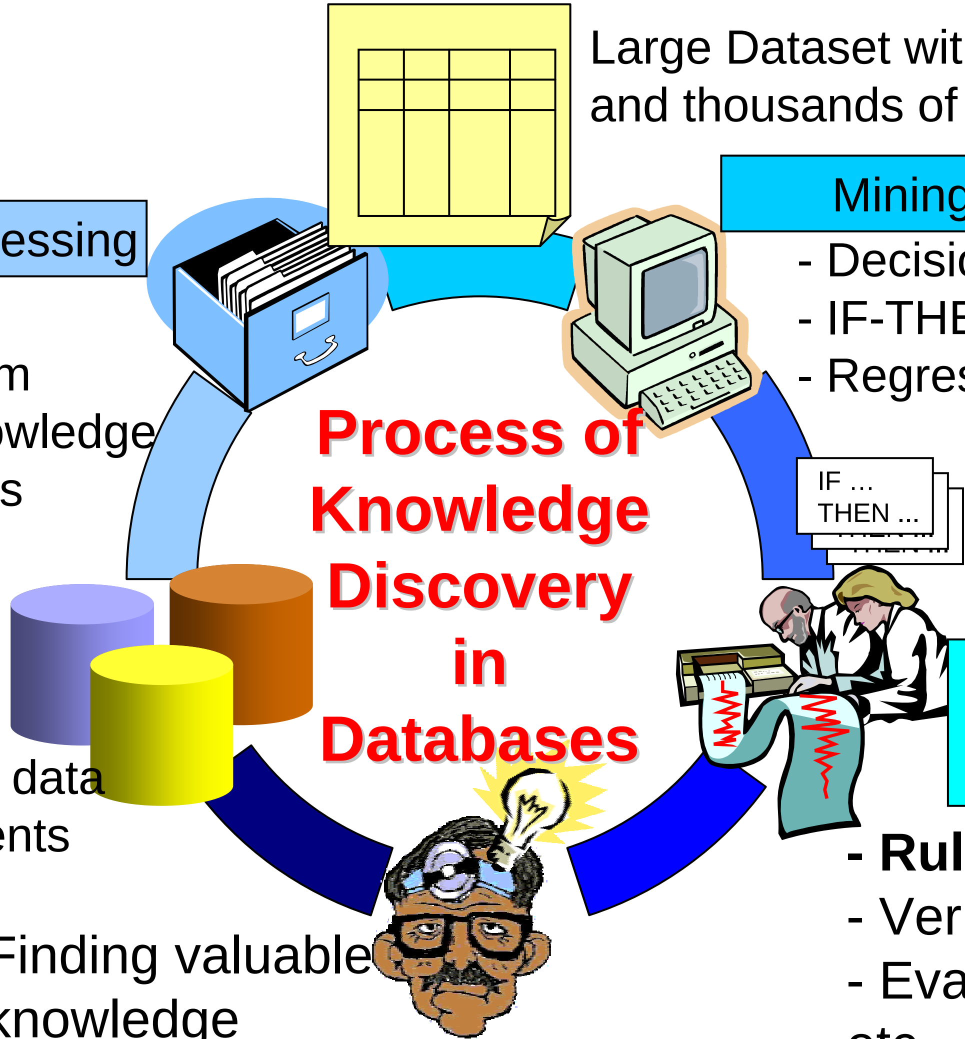
Table of Contents

Table of Contents

- Background
  - datamining process
  - current rule selection problems
  - our research issue
- Rule Evaluation Support Method based on Objective Rule Evaluation Indices
- Comparison of Leaning Algorithms for Rule Evaluation Model Construction
- Conclusion

Background

Large Dataset with hundreds att.
and thousands of inst. including noises



Data Pre-processing

- Data Cleaning
- Adding data from background knowledge
- Coding of values

etc…

Mining

- Decision Tree Learning
- IF-THEN Rule Induction
- Regression models etc…

Many thousands of IF-THEN Rules

Databases

- laboratory test data
- history of patients

etc…

Finding valuable knowledge

Post-processing of Results

- **Rule Selection**
- Verification
- Evaluation

etc…

It is difficult for human experts to evaluate large number of rules completely!!

Background about rule selection

- Many efforts have done to select rules with single objective index such as recall, precision, and so forth.
- At least 40 objective interestingness measures
  - Based on rule evaluation indices
  - Based on similarities between rules

• Ohsaki et al. investigated the relationship between each index and criterion of an expert. However, no single objective index can express the human criterion exactly. [Ohsaki04].
• The availability of each interesting measure never validated on other domains excepting the domain on their paper

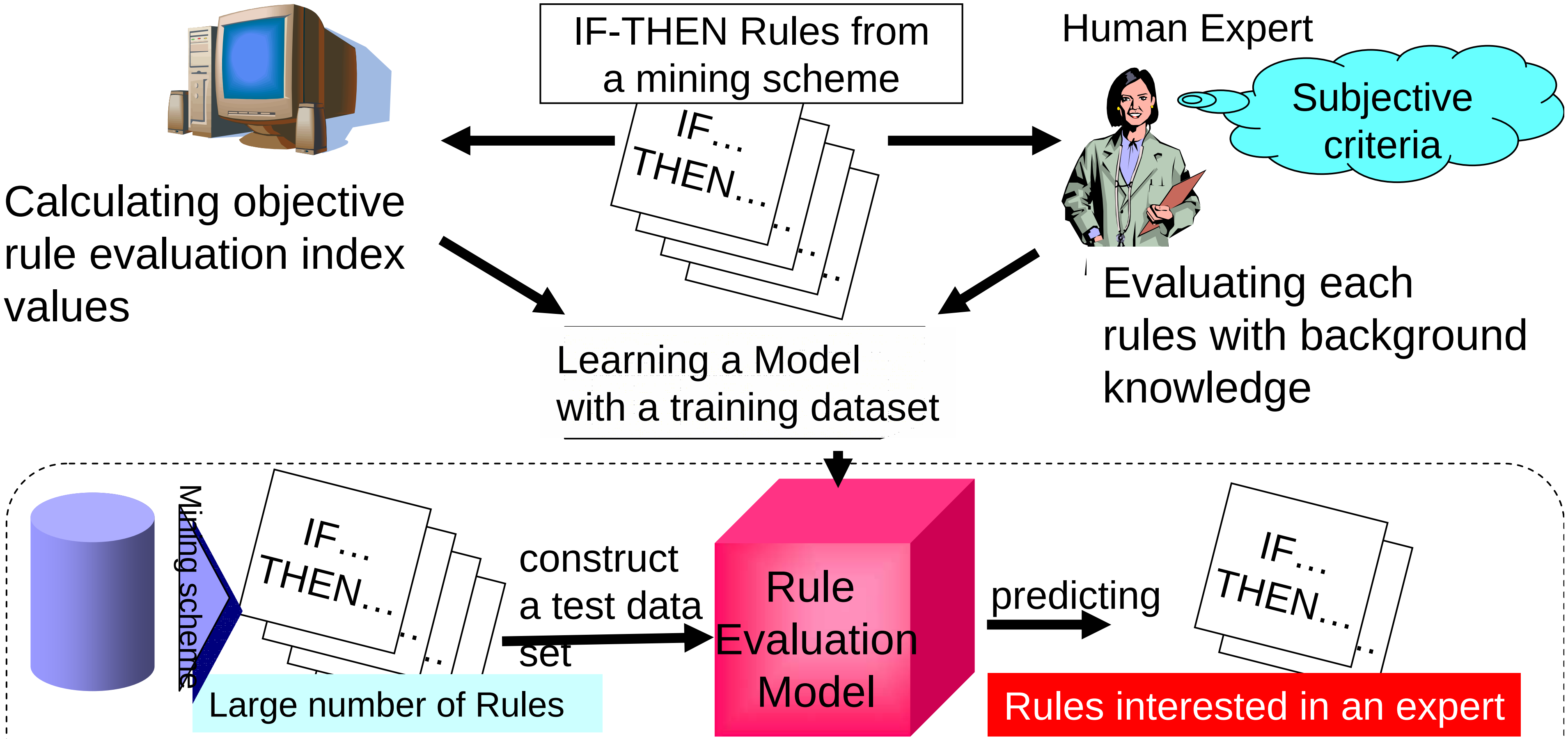
Research Issue

- It is difficult to predict a human criterion with single objective index.
  - We construct rule evaluation models (REMs), combining multiple objective indices and evaluations of a human expert.
- With REMs, we have implemented rule evaluation support method.
  - The system need accurate REMs to support human experts more exactly.

Table of Contents

- Background
- Rule Evaluation Support Method based on Objective Rule Evaluation Indices
- Comparison of Leaning Algorithms for Rule Evaluation Model Construction
- Conclusion

Overview of rule evaluation support with rule evaluation models



Choosing valuable rules with predicting evaluation labels based on values of rule evaluation indices of the given rules

Table of Contents

A comparison of learning algorithms to construct rule evaluation models

- To construct more accurate REMs to support a human expert more exactly
  - Obtaining a data set consists of objective rule evaluation indices and human evaluations

- Viewpoints of this comparison:
  - Accuracies to the whole dataset and Leave-One-Out validation, and their recalls and precisions of each class label
  - Minimum size of training subset to construct valid REMs
  - Contents of REMs from an actual data mining result

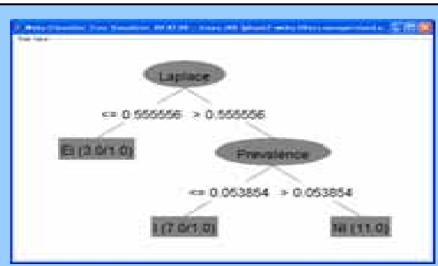
Overview of the case study with the meningitis datamining result

meningitis datasets having 140 inst.
6 (2×3) kinds of diagnostic problems

244 rules

39 objective indices as attributes of the training dataset

Evaluations from a medical expert as class labels



- decision tree
- neural network
-support vector machine
-etc..

Model Construction

Evaluation on:
- The whole training dataset
- Leave-one-out
- Learning curves

Sample of the data set

ruleID	Accuracy	Added_Value	・・・	YulesQ	YulesY	HumanExpert
Rule1	0.81	0.41	・・・	0.73	0.44	NI
Rule10	0.81	0.43	・・・	0.75	0.45	NI
Rule11	0.85	0.46	・・・	0.79	0.49	I
Rule12	0.84	0.56	・・・	0.87	0.58	I
Rule13	0.94	0.44	・・・	0.88	0.59	I
Rule14	0.81	0.43	・・・	0.75	0.45	NI

← 39 objective rule evaluation indices →

Objective Rule Evaluation Indices

calculated on a validation dataset for each IF-THEN rule

The 39 objective indices [Ohsaki 04]

- Based on probability (26 indices)
  - Coverage, Prevalence, Precision, Recall, Support, Accuracy, Specificity, Lift, Leverage, Added Value, Relative Risk, Jaccard, Certainty Factor, Odds ratio, Yule's Q, Yule's Y, Kappa, Koelesgen's Interestingness, Brin's Interestingness,Brin's Conviction, GOI, Credibility, KSI, Laplace Correction, Collective Strength
- Based on test statistics (3 indices)
  - Chi-Square(with only True/Positive, with a whole confusion matrix) , Gini Gain
- Based on information theory (6 indices)
  - Mutual Information, J-Measure, YLI1, YLI2, YZI, K-Measure
- Based on number of instances (3 indices)
  - **Φ** coefficient, PSI , Cosine Similarity
- Based on similarity between rules on a validation dataset (2 indices)
  - GBI, Peculiarity

Learning algorithms for comparison

- Decision Tree: J4.8 (an Java implementation of C4.5)
- Neural Network: BPNN (with back-propagation)
  - Parameters of BP： learning rate=0.3， momentum= 0.2
  - Each unit corresponds to each class label in output layer
- Classification Via Linear Regression: CLR
  - Linear regression expressions: “1-the other” for each class label
  - explanatory variable selection: greedy search with AIC
- Sequential Minimal Optimization [Platt98]: SVM
  - SVM for multiple class: learning “1-the other” expressions for each class label
  - Kernel function setting: polynomial kernel
- OneR
  1. sorting with just one att.
  2. setting thresholds based on class labels
  3. constructs a rule set

Performance Comparison of the five algorithms

(All of rules =244, 'I'=48(19.7%), 'NI'=187(76.6%), 'NU'=9(3.7%))

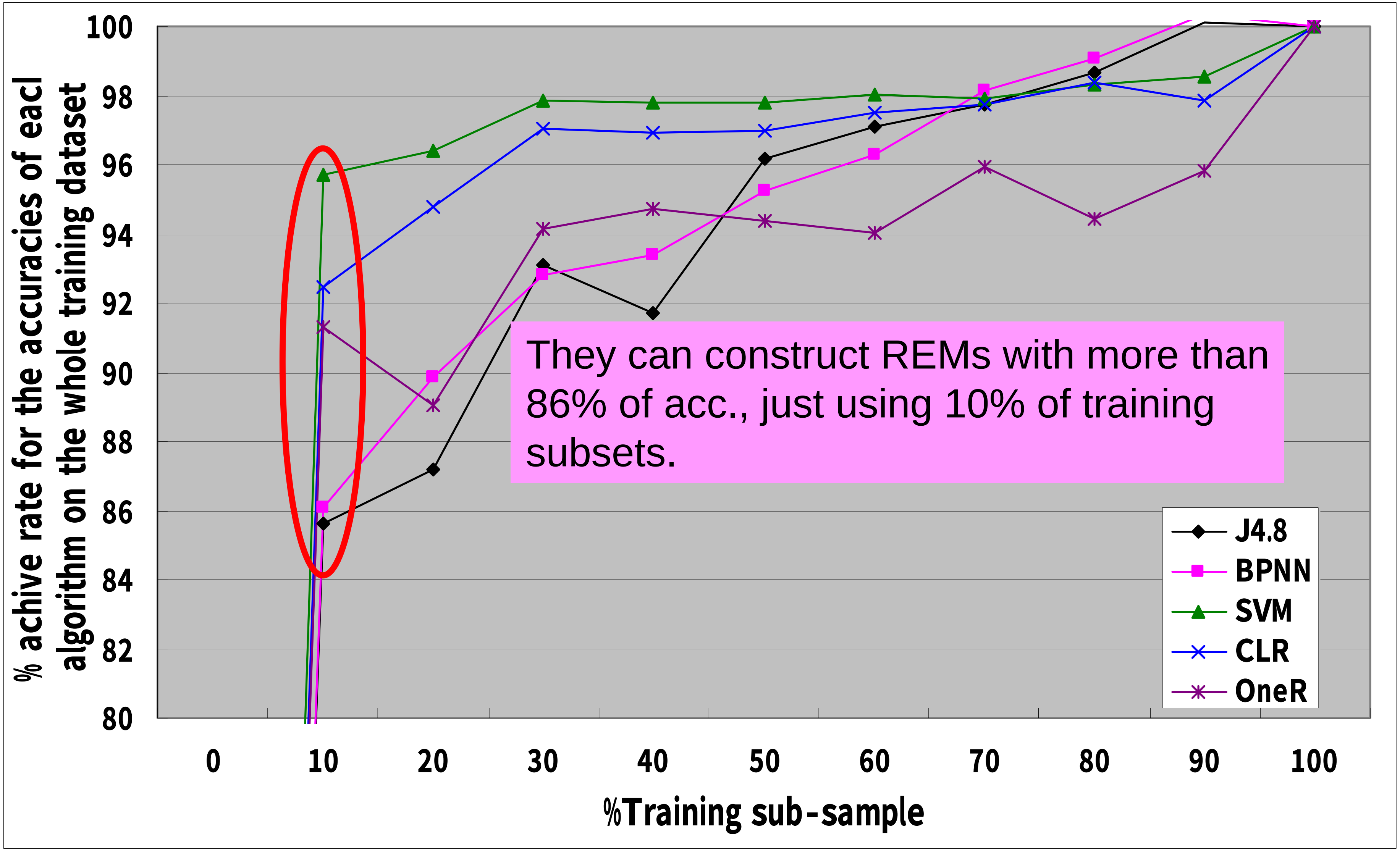
Learning Algorithms	Evaluation on the Whole Training Dataset						
	Acc.	Recall			Precision		
		I	NI	NU	I	NI	NU
J4.8	85.7	41.7	97.9	66.7	80.0	86.3	85.7
BPNN	86.9	81.3	89.8	55.6	65.0	94.9	71.4
SVM	81.6	35.4	97.3	0.0	68.0	83.5	0.0
CLR	82.8	41.7	97.3	0.0	71.4	84.3	0.0
OneR	82.0	56.3	92.5	0.0	57.4	87.8	0.0

Learning Algorithms	Evaluation with Leave－One-Out(LOO)						
	Acc.	Recall			Precision		
		I	NI	NU	I	NI	NU
J4.8	79.1	29.2	95.7	0.0	63.6	82.5	0.0
BPNN	77.5	39.6	90.9	0.0	50.0	85.9	0.0
SVM	81.6	35.4	97.3	0.0	68.0	83.5	0.0
CLR	80.3	35.4	95.7	0.0	60.7	82.9	0.0
OneR	75.8	27.1	92.0	0.0	37.1	82.3	0.0

1. BPNN and J4.8 achieve higher than 85.7% of acc. with more than 77.5% reliability.
   (BPNN tend to be over fitting, looking at it's LOO acc., recalls and precisions)
2. 'NU' is difficult to predict, because of very minor in this rule set.

Leaning curves of accuracies

(achieve rates for the accuracies on the whole training dataset)

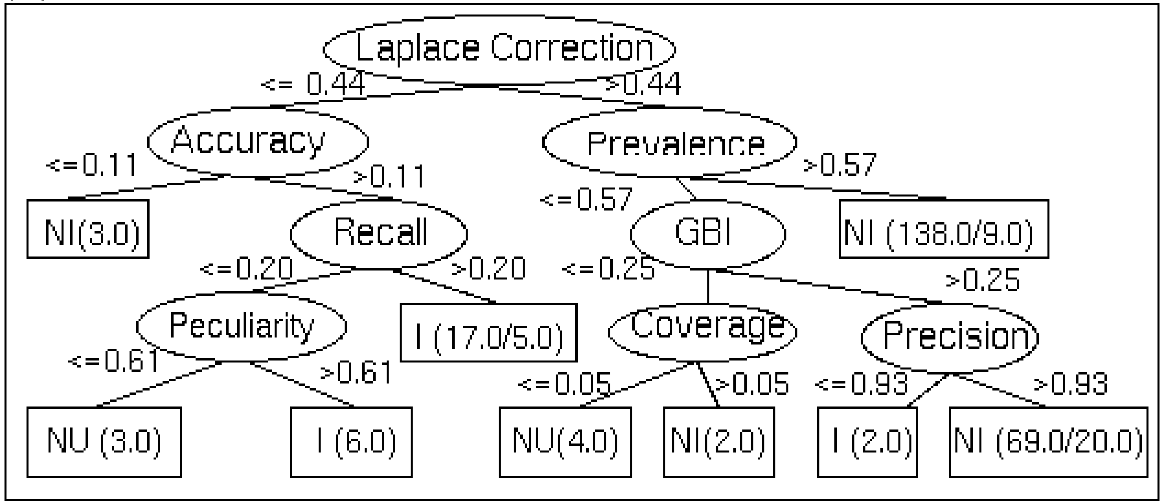


Rule Evaluation Models

(a) The rule set from OneR

```
IF                    YLI1 <  0.02 THEN "I"
IF YLI1 >= 0.02 and YLI1 <  0.29 THEN "NI"
IF YLI1 >= 0.29 and YLI1 <  0.43 THEN "I"
IF YLI1 >= 0.43 and YLI1 <  0.44 THEN "NI"
IF YLI1 >= 0.44 and YLI1 <  0.55 THEN "I"
IF YLI1 >= 0.55 and YLI1 <  0.63 THEN "NI"
IF YLI1 >= 0.63 and YLI1 <  0.83 THEN "I"
IF YLI1 >= 0.83                  THEN "NI"
```

(b) The decision tree from J4.8



(c) The linear regression expressions from CLR

```
NU =
   0.6202 * Specificity +
   0.6224 * Accuracy +
  -1.1384 * Leverage +
  -0.6895 * RelativeRisk +
   0.3704 * CertaintyFactor +
   0.5722 * OddsRatio +
   0.7656 * BI +
  -0.222  * Credibility +
  -0.3941 * LaplaceCorrection +
   0.7986 * GiniGain +
  -0.0966 * GBI +
  -0.8895

NI =
   1.7173 * Precision +
  -0.5063 * Accuracy +
   0.5673 * RelativeRisk +
  -1.2718 * CertaintyFactor +
   0.5955 * YulesQ +
  -0.4609 * K-Measure +
   0.4613 * PSI +
  -0.4181 * Peculiarity +
   0.5302

I =
  -1.4417 * Precision +
  -0.7286 * Specificity +
   0.4085 * Lift +
   0.6297 * CertaintyFactor +
  -1.4477 * CollectiveStrength +
   1.5449 * GiniGain +
  -0.5318 * PSI +
   0.4981 * Peculiarity +
   1.4872
```

Table of Contents

- Background
- Rule Evaluation Support Method based on Objective Rule Evaluation Indices
- Comparison of Leaning Algorithms for Rule Evaluation Model Construction
- Conclusion

Conclusion

- Summary
  - Comparing learning algorithms to construct rule evaluation models for supporting a post-processing of data mining exactly
    - We have achieved that the algorithms can construct accurate rule evaluation model with 39 objective rule evaluation indices.
    - The algorithms have been able to construct a valid rule evaluation model with 10% of training subset.
    - The algorithms constructed each model with different objective indices.

- Feature works
  - Algorithm selection support for rule evaluation model construction
    - attribute construction and selection algorithm selection
    - model learning algorithm selection
  - Applying this method to other data from other domains