TECHNISCHE UNIVERSITÄT DARMSTADT

Knowledge Engineering Group

An Empirical Quest for optimal Rule Learning Heuristics

Setup

Simple Separate-and-conquer algorithm implemented in the *SeCo*-Framework:

- **Conquer Step**: learn a rule from the data (refine until no negative is covered)
- **Separate Step**: remove all examples which are covered by the rule
- Hill-Climbing Search
- different heuristics

Rule Learning Heuristics

Rule Learning Heuristics (*h*) have to optimize 2 criteria simultaneously:

- **Coverage**: maximize number of covered positive examples $h_{cov} = p$
- **Consistency**: minimize number of covered negative examples $h_{con} = -n$

Parametrized heuristics trade off between variants of the 2 criteria:

- *cost-measure* $h_{cost} = c * h_{cov} + (1-c) * h_{con}$
- *relative cost measure* $h_{rcost} = c_r * \frac{p}{P} - (1-c_r) * \frac{n}{N}$
- *m*-estimate $h_{mest} = \frac{p + m * \frac{P}{P+N}}{p+n+m}$
- *Klösgen-measures* $h_{kl} = \frac{p+n}{P+N}^{\omega} * \left(\frac{p}{p+n} - \frac{P}{P+N}\right)$
- *F*-measure $h_{fm} = \frac{(\beta^2+1) * \frac{p}{p+n} * \frac{p}{P}}{\beta^2 * \frac{p}{p+n} * \frac{p}{P}}$

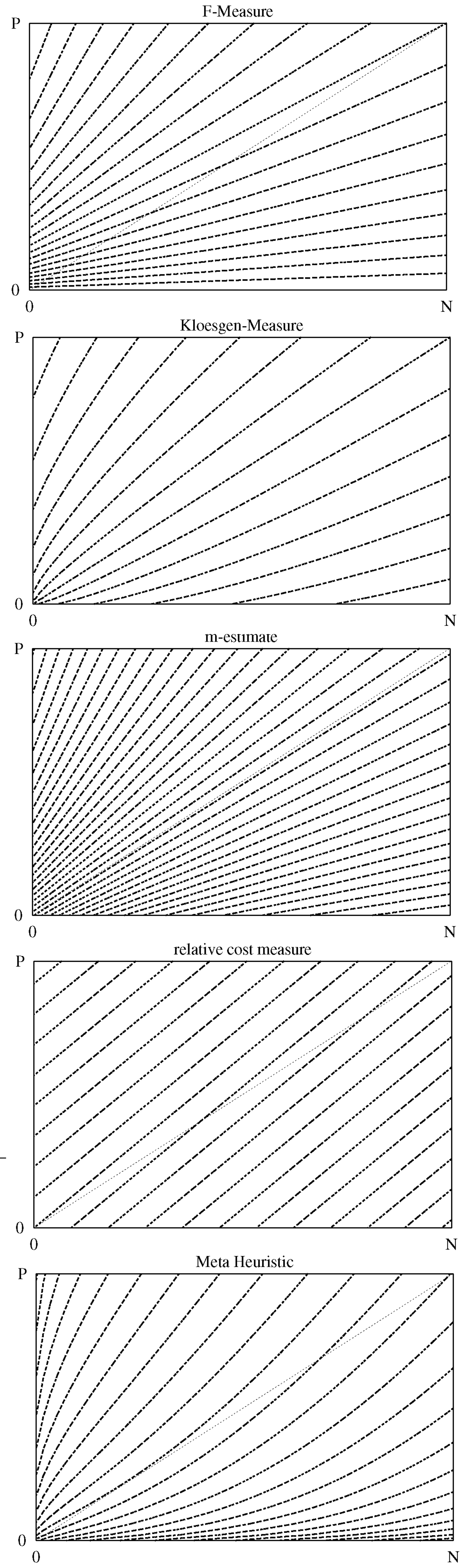
The way to go

1. Optimize the trade-off for the 5 different parametrized heuristics on 27 data sets and test the parametrizations on 30 different sets (all taken from UCI-Repository)
   - start with a set of intuitively appearing parameters (depending on the range)
   - continously narrow down the region of interest
2. Learn a new heuristic from observed rule statistics via meta-learning
   - let the *SeCo*-Algorithm run several times with different heuristics on the 27 sets
   - log statistics of **all** rules (not only final rules)
   - try to fit the meta data (87,380 examples) with a linear regression

Sample Results

- all parametrized heuristics outperform standard heuristics (except the cost-measure)
  - parameters remain stable (Spearman Rank Correlation between Ranking on TuneSets and TestSets was **0.85**)
- *relative cost measure* works best
- *meta heuristic* comparable with *relative cost measure*
  - but only when absolute inputs are logarithmized
- the *a priori* class distribution is necessary to build a good heuristic
- *consistency* should be preferred over *coverage*

heuristic	Macro Avg. Acc.	Size
JRip	78.98	12.20
Meta-Heuristic	78.88	37.03
rel.cost measure	78.87	25.30
m-estimate	78.67	46.33
Klösgen-measure	78.46	61.83
F-Measure	78.12	51.57
Correlation	77.55	47.33
Laplace	76.87	117.00
Consistency	76.22	128.37
cost-measure	76.11	122.87
WRA	75.82	12.00
Accuracy	75.65	99.13



Publications

- Frederik Janssen and Johannes Fürnkranz: *On trading off consistency and coverage in inductive rule learning*. In *K.-D. Althoff and M. Schaaf*, Editors, Proceedings of the LWA 2006, pages 306-313, 2007.
- Frederik Janssen and Johannes Fürnkranz: *On meta-learning rule learning heuristics*. In Proceedings of the 7th IEEE Conference on Data Mining (ICDM-07), pages 529-534, Omaha, NE. 2007.
- Frederik Janssen and Johannes Fürnkranz: *An empirical quest for optimal rule learning heuristics*. Technical Report TUD-KE-2008-01. TU Darmstadt. Knowledge Engineering Group. 2008.

Contact

Technische Universität Darmstadt
Fachbereich Informatik
Fachgebiet Knowledge Engineering
Hochschulstraße 10
D-64289 Darmstadt

Telefon: +49-6151-16-5409
Fax: +49-6151-16-5482
E-mail: ke@informatik.tu-darmstadt.de
http://www.ke.informatik.tu-darmstadt.de