A Re-evaluation of the Over-Searching Phenomenon in Inductive Rule Learning

Frederik Janssen and Johannes Fürnkranz

Motivation
- Phenomenon of over-searching is well known [3] but not shown for most of the rule learning heuristics
- In [3] only one heuristic was used and no true Exhaustive Search was employed (approximation with a beam of 512)
- We extend their work to 9 different heuristics (some of them were tuned in [1,2]) and a true exhaustive search
- We want to answer the question whether Separate-and-conquer algorithms can improve from more extensive search

Setup
- Simple Separate-and-conquer algorithm implemented in the SeCo-Framework
  - Hill-Climbing Search, Exhaustive Search and Beam Search (for a trade-off between them)
  - Implements Forward Pruning (important for the runtime)
  - Classification by decision list (ordered binarization)
- Experiments
  - 22 datasets from UCI (arbitrary selection, only nominal attributes)
  - AV accuracy with 10-fold CV

Search strategies
- **Hill-Climbing**
  - Only refine 1 rule
  - May get stuck in local optima
- **Beam Search**
  - Refine b rules simultaneously
  - Higher runtime
- **Exhaustive Search**
  - Create all possible rules
  - Highest runtime
  - Cannot get stuck in local optima

Results
- Exhaustive Search finds longer rules with higher coverage (cf. Table)
- Experiment 2: Only induce one single rule per class
  - Confirms findings of previous experiment
  - Models only lack about 10% accuracy behind
  - Precision and Laplace have significantly smaller theories
  - All heuristics improve from Exhaustive Search except the Meta-learned one

Discussion
- Over-searching phenomenon depends on the heuristic
  - Odds Ratio and Precision gain performance
  - More complex heuristics lose performance
  - Heuristics that work well in Hill-Climbing do not profit from Exhaustive Search or Beam Search
  - Different requirements for heuristics used in Hill-Climbing and Exhaustive Search

References


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

Phone: +49-6151-16-5409
Fax: +49-6151-16-5482
E-mail: {janssen,juffi}@ke.tu-darmstadt.de
http://www.ke.informatik.tu-darmstadt.de