Workshop on Advances in Inductive Rule Learning

ECML/PKDD 2004
24.9.2004
Separate-and-Conquer Rule Learning

- Learn a set of rules, one by one

1. Start with an empty theory $T$ and training set $E$
2. Learn a single (*consistent*) rule $R$ from $E$ and add it to $T$
3. If $T$ is satisfactory (*complete*), return $T$
4. Else:
   - **Separate**: Remove examples explained by $R$ from $E$
   - **Conquer**: If $E$ is non-empty, goto 2.

- One of the oldest family of learning algorithms
  - goes back AQ (Michalski, 60s)
  - FRINGE, PRISM and CN2: relation to decision trees (80s)
  - popularized in ILP (FOIL and PROGOL, 90s)
  - RIPPER brought in good noise-handling

- Different learners differ in how they find a single rule
<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Language Bias</th>
<th>Search Bias</th>
<th>Overfitting Avoidance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Static</td>
<td>Dyn.</td>
<td>Algorithm</td>
</tr>
<tr>
<td>AQ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AQ15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AQ17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATRIS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BEXA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHAMP</td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>C1PF</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CN2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CN2-MCL</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLASS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DLC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FOCL</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>FOLI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FOSSIL</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>GA-SMART</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>GELE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GREEDY3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GRENDEL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GROW</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HYDRA</td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>LBL-SMART</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>INDUCE</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I-REF, I²-REF</td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>JOG</td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>m-FOIL</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>MDL-FOIL</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>MILP</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ML-SMART</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>NINA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POSEIDON</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PREPEND</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRISM</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PROGOL</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>REP</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>RIPPER</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RDT</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SFOIL</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMART+</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>SWAP-I</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>TDP</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
</tbody>
</table>

- **Language bias:**
  - which type of conditions are allowed (static)
  - which combinations of conditions are allowed (dynamic)

- **Search bias:**
  - search heuristics
  - search algorithm (greedy, stochastic, exhaustive)
  - search strategy (top-down, bottom-up)

- **Overfitting avoidance bias:**
  - pre-pruning (stopping criteria)
  - post-pruning
Rule Evaluation Metrics

• Understanding rule evaluation metrics
  ▪ For classification rules: (Fürnkranz & Flach, MLJ, in press)
  ▪ For rule interestingness: (Freitas, KBS 1999), (Tan et al. KDD-02), (Ohsaki et al., PKDD-04)
  ▪ Unifying view: (Lavrac, Flach, Zupan, ILP-99)

• Other approaches
  ▪ Patient Rule Induction (Friedman & Fischer, 1999)
  ▪ Using ROC convex hull (Prati & Flach, today)
  ▪ Meta-Learning of heuristics (Fürnkranz, today)

• Search Heuristics ≠ Evaluation Metrics
  ▪ heuristics for picking the best refinements need not be the same as the heuristics for picking the best rule
Noise Handling

- **Pre-Pruning:**
  - various heuristics and stopping criteria
  - attempt for analytical comparison (Fürnkranz & Flach, ECML-04)
  - still room for improvement

- **Post Pruning:**
  - Incremental Reduced Error Pruning (Fürnkranz & Widmer, ICML-94)
  - I-REP for unordered rule sets (Boström, today)
Handling Multiple Classes

- Conventional approaches (both in CN2)
  - **decision lists**: class label is chosen after the rule has been learned, for learning decision lists
  - unordered rule induction: learn a theory for each class
  - comparison of both approaches (van Zyl & Cloete, ECML-04)

- Ordered Induction (Ripper)
  - like unordered, but order classes first (e.g., frequency)

- Round robin learning / pairwise classification
  - learn a theory for each pair of classes (Fürnkranz, JMLR 02)
Combining Rules

- Conventional approach:
  - pick first rule that fires
  - pick rule with highest Laplace-corrected precision
- Improving Default Classification
  - Rule Stretching (Eineborg & Boström, ILP-01)
- Meta-Learning for Conflicts
  - Double Induction (Lindgren & Boström, IDA-03)
  - Recursive Induction (Lindgren, ECML-04)
- Meta-Learning for Conflicts & Default Classification
  - (Ramirez & Hazan, today)
- For optimizing AUC performance
  - weighted voting, lowest FPR (Fawcett, ICDM-01)
Dealing with Redundant Rule Sets

- Association Rules
  - CBA (Liu et al., KDD-98), APRIORI-C (Jovanovski & Lavrac, 2001)
- Stochastic Local Search
  - (Rückert & Kramer, ICML-03, ICML-04)
- Rule Ensembles
  - SLIPPER (Cohen & Singer, AAAI-99), LRI (Weiss & Indurkhya, ICML 2000)
Increasing Expressiveness

- Relational Learning, ILP
- Disjunctions
  - Internal Disjunctions (Michalski, 1980; Theron & Cloete 1996)
- Fuzzy Conditions
  - (van Zyl & Cloete, today)
- Templates, Schemas, and Macros
  - much work in ILP, (Pena Castillo & Wrobel, today)
- Rules with Constraints
  - Ripple-Down Rules (Scheffer, ca. 1993)
  - Exception Rules (Suzuki, 1997 – today)
- Using Alternative Classifiers
  - Delegating Classifiers (Cerri et al., ICML-04)
- Multi-Instance Rule Learning (Chevaleyre & Zucker, ECML-01)
Feature Engineering

- Rule-Based Feature Construction
  - Stochastic Propositionalisation (Kramer et al., ILP-98)
  - (van den Bosch, today)
- Feature Subset Selection and Feature Relevancy
  - (Lavrac & Gamberger, 1998 - today)
Alternatives to Covering

- Weighted Covering
  - SLIPPER (Cohen & Singer, AAAI-99), LRI (Weiss & Indurkhya, ICML 2000), (Kavsek & Lavrac ← today)
- Bottom-up Generalization of Rule Sets
  - RISE (Domingos, MLJ 96), PLCG (Widmer, AIJ 02), SUNRISE (de Pina & Zaverucha ← today)
- Refinement Operators for Rule Sets
  - HYDRA (Bratko, ILP-99), (Badea, ILP-01), INTHELEX (Esposito et al., MLJ 2000)
- Randomized Rule Learning
  - SLS (Rückert & Kramer, ICML-04), (Pfahringer et al. ← today)
- Incremental Rule Learning
  - various works in theory refinement, no practical system yet
Learning Theory

- Set Covering Machine
  - (Marchand & Shawe-Taylor, JMLR 02)
- Decision List Machine
  - (Sokolova, Marchand, Japkowicz, Shawe-Taylor, NIPS-03)
Beyond Classification

- Regression
  - FORS (Karalic & Bratko, MLJ 1993), SWAP1R (Weiss & Indurkhya, JAIR 95), R2 (Torgo, 1995), Cubist

- Ordinal Regression
  - (Horvath & Vojtas, *today*)

- Clustering
  - CLUSTER/2 (Michalski & Stepp 1983), Some work in ILP, (Pelleg & Moore, ICML-01), (Mishra et al., MLJ 2004)

- Preference Learning
  - (Fürnkranz & Hüllermeier, ECML-03)

- Subgroup Discovery
  - (Lavrac et al., JMLR-04, MLJ-04, *today*)
The list of topics and works on the previous pages is necessarily incomplete. Please let me know of anything important that is missing (like your own work 😊).

juffi@ke.informatik.tu-darmstadt.de