Krzysztof Grąbczewski

Meta-Learning in Decision Tree Induction
## Contents

1 Introduction ............................................. 1  
  1.1 Learning Machines and Meta-Learning .................. 2  
  1.2 Basic Definitions and Notations ..................... 4  

2 Techniques of Decision Tree Induction ........................ 11  
  2.1 Recursive Top-Down Splits ............................ 12  
  2.2 Univariante Decision Trees ......................... 14  
    2.2.1 ID3 .............................................. 14  
    2.2.2 CART ........................................... 16  
    2.2.3 C4.5 ............................................ 18  
    2.2.4 Cal5 ........................................... 20  
    2.2.5 FACT, QUEST and CRUISE ........................ 23  
    2.2.6 CTree .......................................... 31  
    2.2.7 SSV ............................................. 34  
    2.2.8 ROC-Based Trees ................................ 36  
  2.3 Multivariate Decision Trees ........................... 39  
    2.3.1 LMDT ............................................ 40  
    2.3.2 OC1 ............................................. 43  
    2.3.3 LTree, QTree and LgTree ........................ 45  
    2.3.4 DT-SE, DT-SEP and DT-SEPIR .................... 48  
    2.3.5 LDT ............................................. 50  
    2.3.6 Dipolar Criteria for DT Induction ............... 53  
  2.4 Generalization Capabilities of Decision Trees ............ 55  
    2.4.1 Stop Criteria ................................... 56  
    2.4.2 Direct Pruning Methods ........................ 57  
    2.4.3 Validation Based Pruning ......................... 63  
  2.5 Search Methods for Decision Tree Induction .............. 70  
  2.6 Decision Making with Tree Structures .................. 77  
  2.7 Unbiased Feature Selection ........................... 79  
  2.8 Ensembles of Decision Trees .......................... 86  
    2.8.1 Option Decision Trees .......................... 88  
    2.8.2 Bagging and Wagging ............................ 89  
    2.8.3 Boosting ....................................... 91
5 Meta-Level Analysis of Decision Tree Induction

5.1 Results Comparison Techniques
5.1.1 Bad Testing Practices
5.1.2 Reliable and Just Comparisons

5.2 Test Scenarios for DT Induction Analyses

5.3 Single Decision Tree Models
5.3.1 Algorithms
5.3.2 Experiment
5.3.3 General Results Visualization and Analysis
5.3.4 Analysis of Results Subgroups
5.3.5 Summary

5.4 Cross-Validation Committees
5.4.1 DTCV Committee Algorithm
5.4.2 Experiment
5.4.3 Win Counts
5.4.4 DTCV Committees Versus Single Validated Trees
5.4.5 DTCV Committees Versus Bagging and Boosting
5.4.6 Algorithm Parameters Analysis
5.4.7 Summary

References

6 Meta-Learning

6.1 Meta-Learning Approaches
6.1.1 No Free Lunch Theorems
6.1.2 Ensembles of Decision Models
6.1.3 Meta-Level Regression
6.1.4 Rankings of Algorithms
6.1.5 Meta-Learning as Active Search

6.2 Meta-Learning as Search with Feedback from Validation
6.2.1 The Algorithm
6.2.2 Proper Meta-Learners
6.2.3 Task Requests and Task Running

6.3 Meta-Learning with Configuration Generators
and Complexity Control
6.3.1 CDML as an Instance of GML
6.3.2 Machine Configuration Generators
6.3.3 Complexity Control
6.3.4 Analysis of Finished Tasks and the Quarantine
6.3.5 Machine Complexity Evaluation
6.3.6 Learning Evaluators
6.3.7 Example Experiment

6.4 Profile-Based Meta-Learning
6.4.1 The Algorithm
6.4.2 Profile Management
7 Future Perspectives of Meta-Learning

Appendix A.

Appendix B.

Index