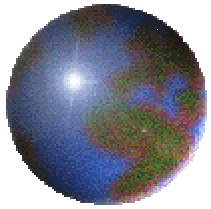
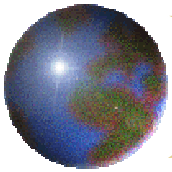


Finding Optimal Decision Trees

Research-in-progress (look into prepared PhD Thesis)

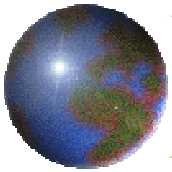
Petr Masa

October 2005



Outline

- Introduction to Decision Trees
- Problem definition
- Terms
- Finding Optimal Decision Trees Theorem
- Algorithm for Real Data
- Results of Algorithm Tests
- Conclusions



Introduction to Decision Trees

- Greedy algorithm
- Post-pruning

- May not fit the distribution
- More Trees can describe the same distribution

Introduction To Decision Trees

- Restriction To Binary Variables

- Theoretical Results: Distribution
- Practical Results: Finite Data

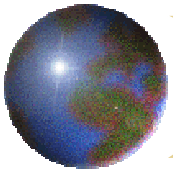
Problem Definition

- Example
- First variable can be “greedily” chosen incorrectly
- Prune is not able to fix it in almost all cases
- Simple = easy understandable, helps to prevent overfitting

Terms

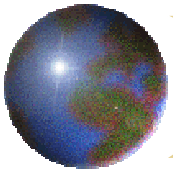
- Strong Faithfulness
- Optimal Decision Tree

- Prune Operation
- Parent-Child Exchange Operation



Finding Optimal Decision Trees Theorem

- For strong Faithful distribution
  - Non strong Faithful is measure zero
- Optimal Decision Tree can be found by sequence of Prune and PCE
  - Sequence may be first only PCEs, then only prunes
- There is only one Optimal Decision Tree



Algorithm

- Developed algorithm to work (=not no hang) on any distribution
- Combines prunes and PCEs

- Polynomial in Number of Leaves of Tree from Greedy Phase
  - Exhaustive search is exponential

Results of Algorithm Tests

- Tests on simulated data

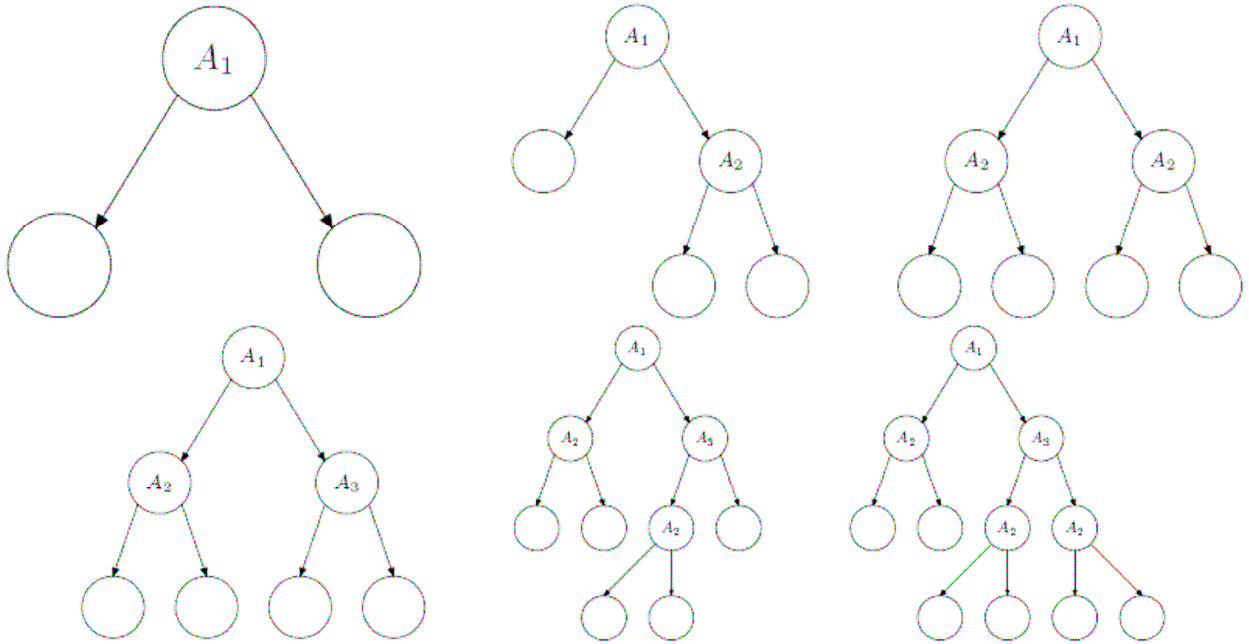
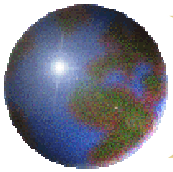


Figure 4.1: Patterns tested – Pattern A, Pattern 1, Pattern B, Pattern 2, Pattern 3 and Pattern 4



Results of Algorithm Tests

- Tests on simulated data

Train set/test set size	Pattern A		Pattern B		Pattern 1		Pattern 1b		Pattern 2	
	PO	PP	PO	PP	PO	PP	PO	PP	PO	PP
300/100	7	7	5	3	2	2	3	3	3	4
1000/300	7	7	6	5	3	5	5	5	4	4
3000/1000	8	8	8	8	4	7	5	8	4	7
7500/2000	8	8	8	8	5	7	5	8	4	7
15000/5000	8	8	8	8	4	7	5	8	4	8
20000/6500	8	8	8	8	5	8	5	8	4	8
30000/10000	8	8	8	8	5	8	5	8	4	7

Results of Algorithm Tests

- Tests on simulated data

Train set/test set size	Pattern 2b		Pattern 3		Pattern 3b		Pattern 4		Pattern 4b	
	PO	PP	PO	PP	PO	PP	PO	PP	PO	PP
300/100	3	2	0	3	1	1	0	0	0	0
1000/300	4	2	1	4	2	5	2	1	3	3
3000/1000	2	5	2	6	1	3	2	2	2	3
7500/2000	4	5	1	8	2	7	3	5	4	5
15000/5000	4	5	1	8	2	7	3	5	4	4
20000/6500	5	7	1	7	1	7	3	5	4	6
30000/10000	5	7	1	6	2	7	3	6	4	6

Results of Algorithm Tests

- Test on Real Data

- CART: 5 leaves, CART-PP: 4 leaves
- Different attribute in the root

Conclusions

- Problem Defined
- Terms introduced
- Theoretical Results Shown
- Algorithm Introduced
- Results on Simulated and Real Data Shown