SHIFTTREE:
AN INTERPRETABLE MODEL-BASED APPROACH FOR TIME SERIES CLASSIFICATION

**Balázs Hidasi**   **- hidasi@tmit.bme.hu**

Csaba Gáspár-Papanek - gaspar@tmit.bme.hu

Budapest University of Technology and Ecomonics,

Hungary

OUTLINE

- General concept
- Task & adaptation
- Labeling & learning
- Evaluation
- Forest methods
- Conclusion

GENERAL CONCEPT

- Not limited to time series
  - Time series
  - Semi-structured data
  - Graphs
- Two questions:
  - „Where to look at?"
  - „What to observe?"
- Operator families
  - EyeShifter Operator(s) (ESO)
  - ConditionBuilder Operators (CBO)
- Dynamic attributes
- Model: sequence of simple operatos / rules

TIME SERIES CLASSIFICATION
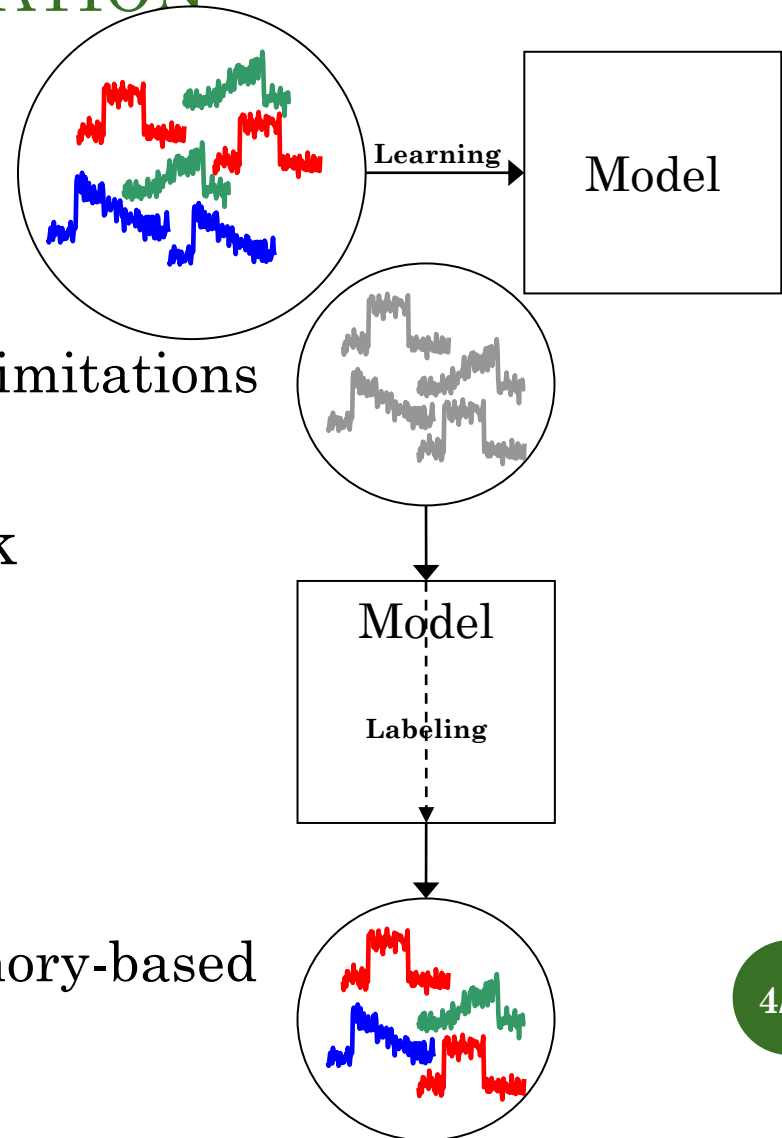
- Time series
  - With one observed variable
  - Evenly sampled
  - The algorithm has no such limitations

- Standard classification task
  - Focusing on accuracy

- Goals
  - Satisfying accuracy
  - Model-based instead of memory-based
  - Interpretable models

Learning

Model

Model

Labeling

Explaining the goals

- Why model-based?
  - Labeling is much faster ✓
  - Can learn more general properties of the classes ✓
  - Needs more training samples ✗

- What is interpretability good for?
  - Getting user trust ✓
  - Helping in understanding the data ✓

CONCEPT APPLIED TO TIME SERIES

- Dynamic attributes of time series
  - EyeShifter Operator (ESO): „Where to look at?"
    - Moves a cursor along the time axis
    - To a specified point of the series
    - E.g.: „To the next local minimum", „100 steps ahead", etc.
    - Note: result of operator sequence depens on the order
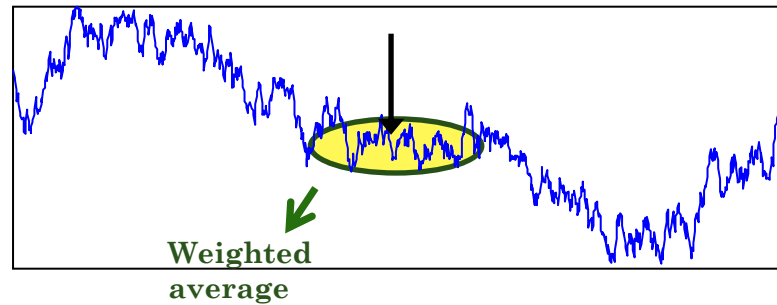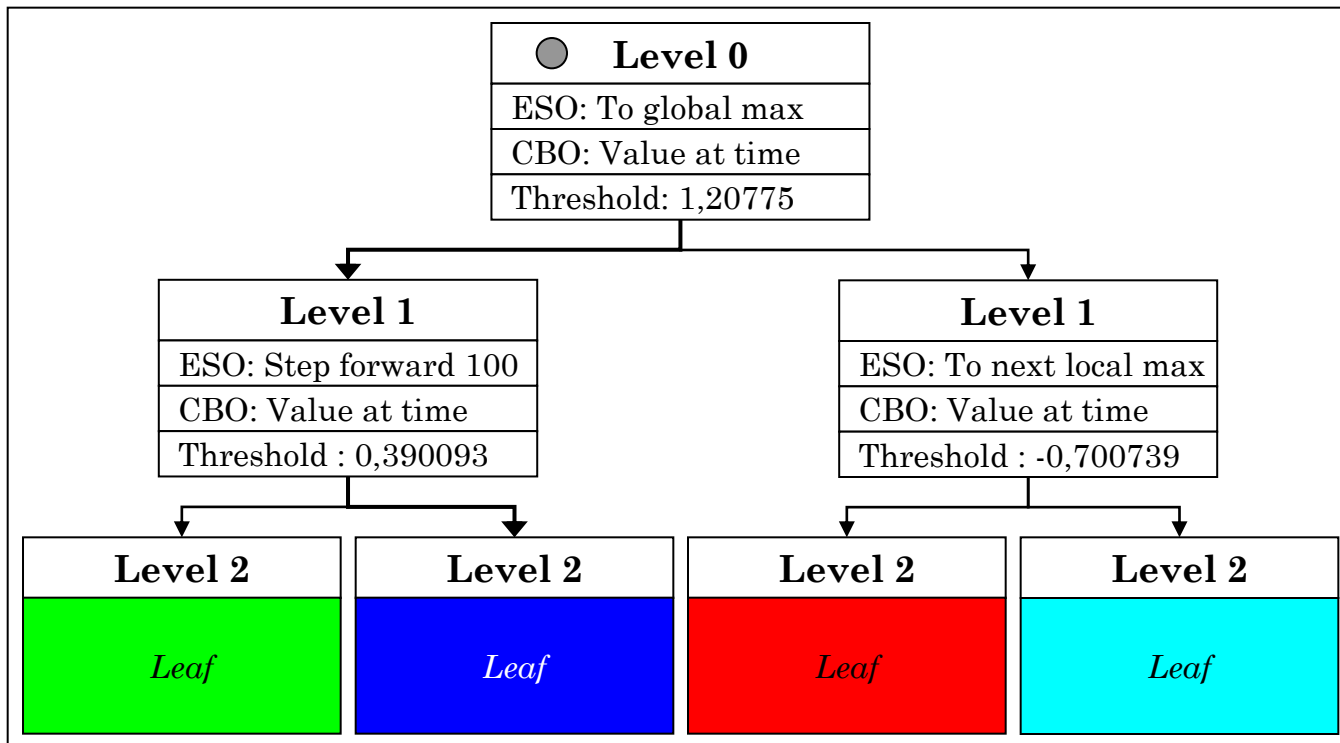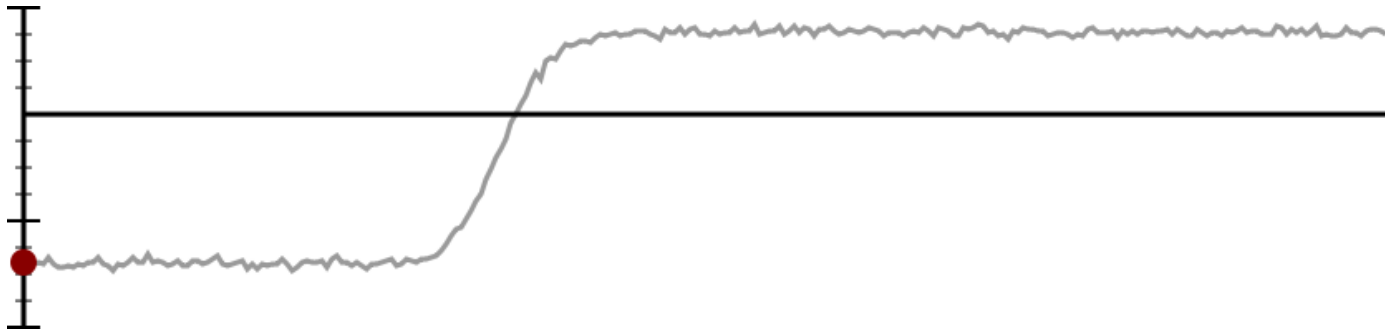 - ConditionBuilder Operator (CBO): „What to observe?"
    - Computes the dynamic attribute
    - E.g.: „Value at the given time", „Length of the jump", etc.

Weighted average

- Decision tree as model
  - Works well with the dynamic attribute concept
  - Operator sequence from root to leaf

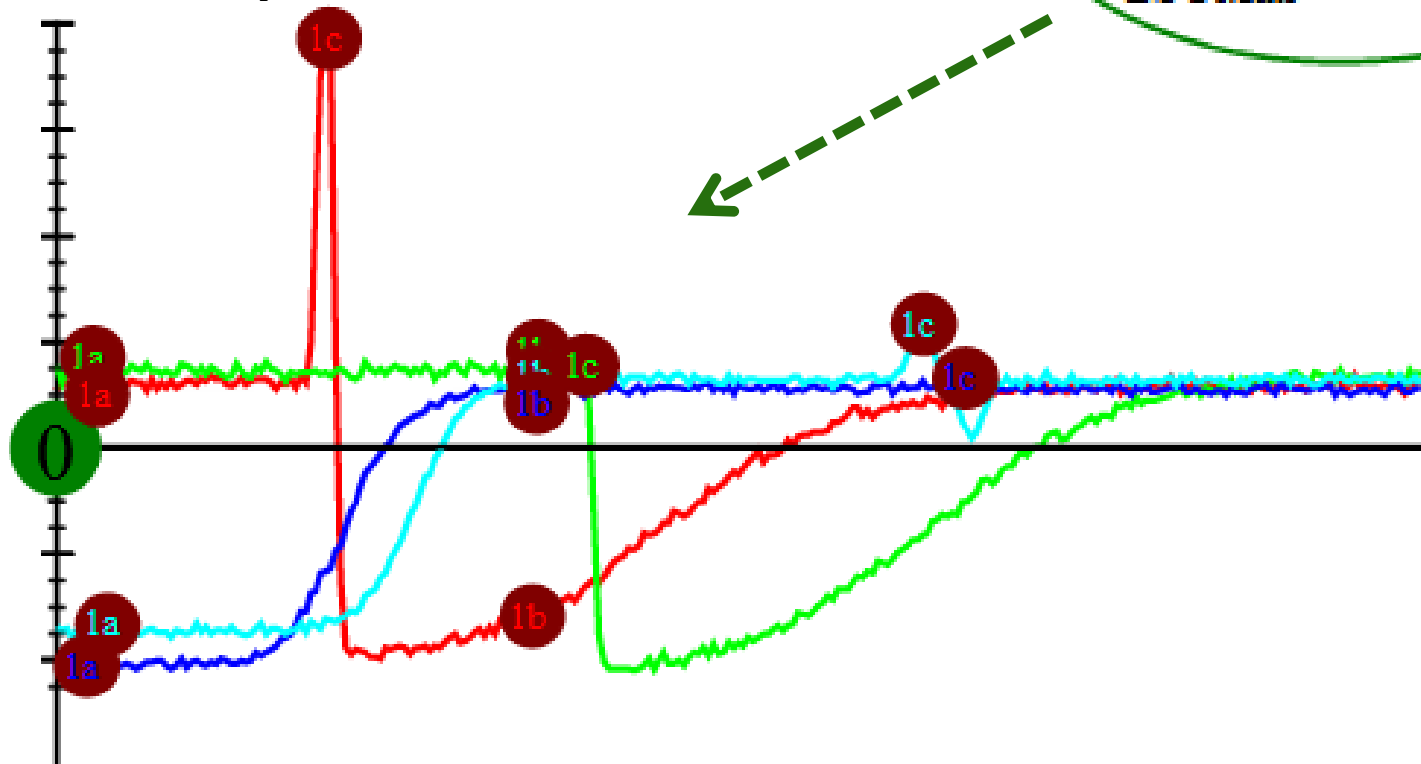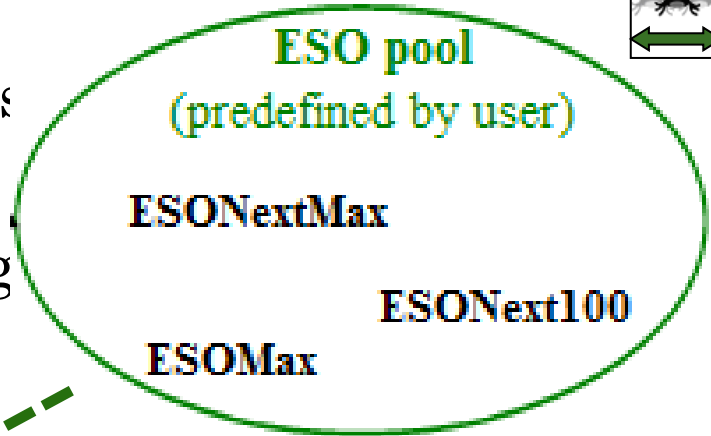# LABELING

LEARNING (1/3)

- Operator pool definition: ESO
- Root node
- Cursor starts at the beginning
  - May be overridden

ESO pool
(predefined by user)

ESONextMax

ESONext100

ESOMax

LEARNING (2/3)

CBO pool
(predefined by user)

CBOSimple

CBOAvg20

- Available dynamic attributes (CBO pool)
  - At/around reachable positions
- Selecting a…
  - Attribute
  - Threshold value

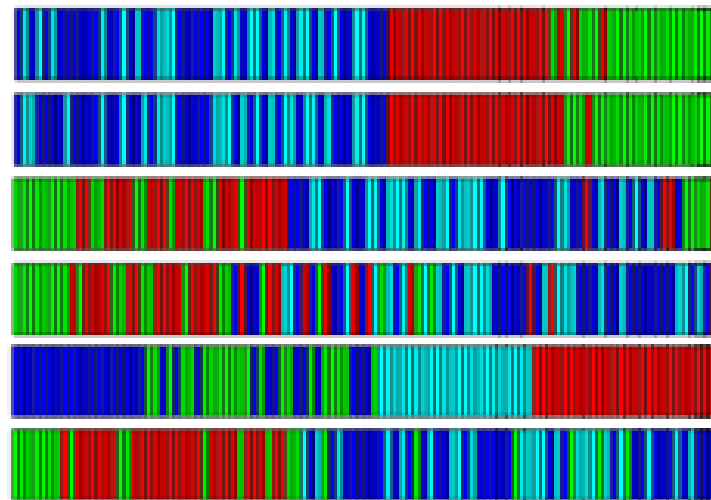Minimize weighted child node entropy
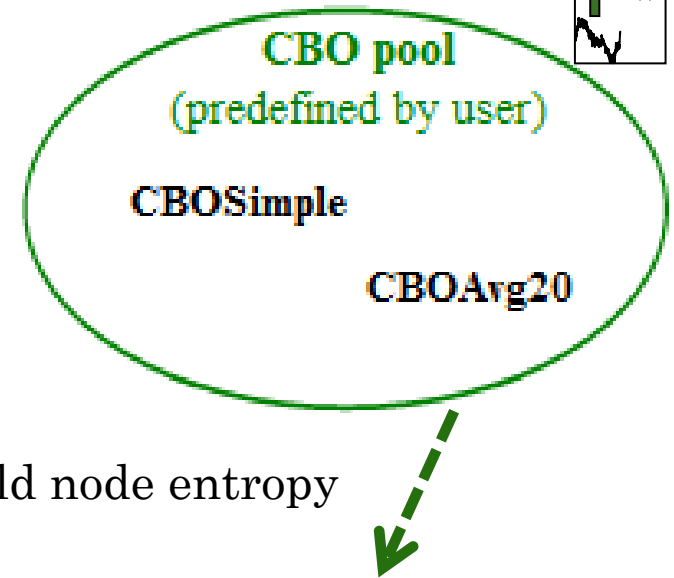
Possible dynamic attributes
(operator pairs) in root node
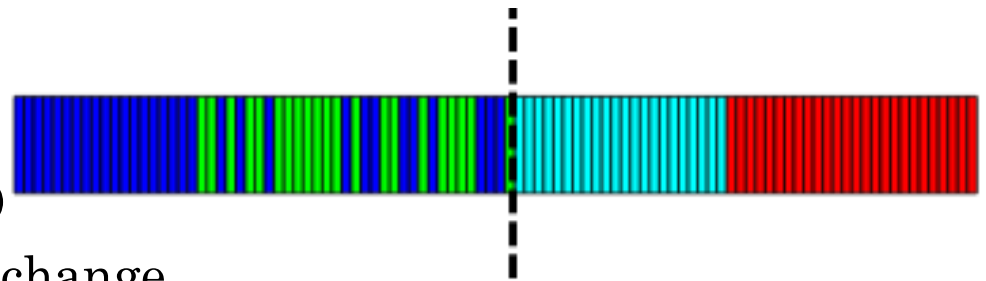
Ordering of the samples in the node
by dynamic attributes

Value at **1a**

Average around **1a** using R = 20

Value at **1b**

Average around **1b** using R = 20

Value at **1c**
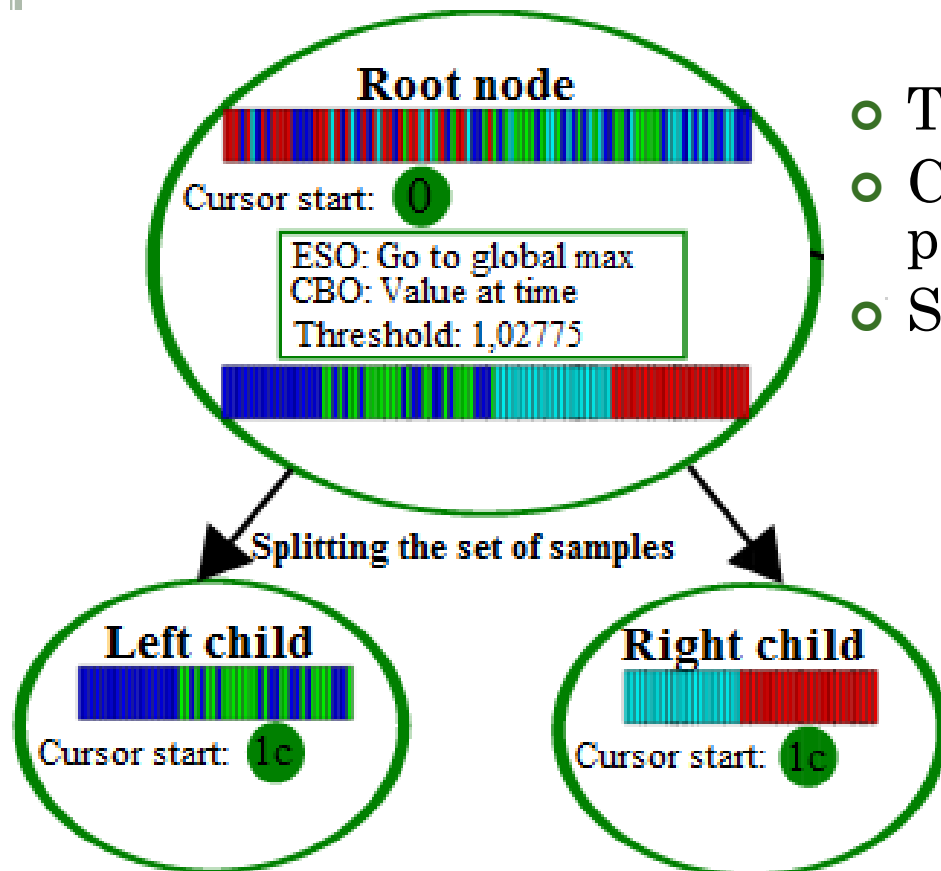
Average around **1c** using R = 20
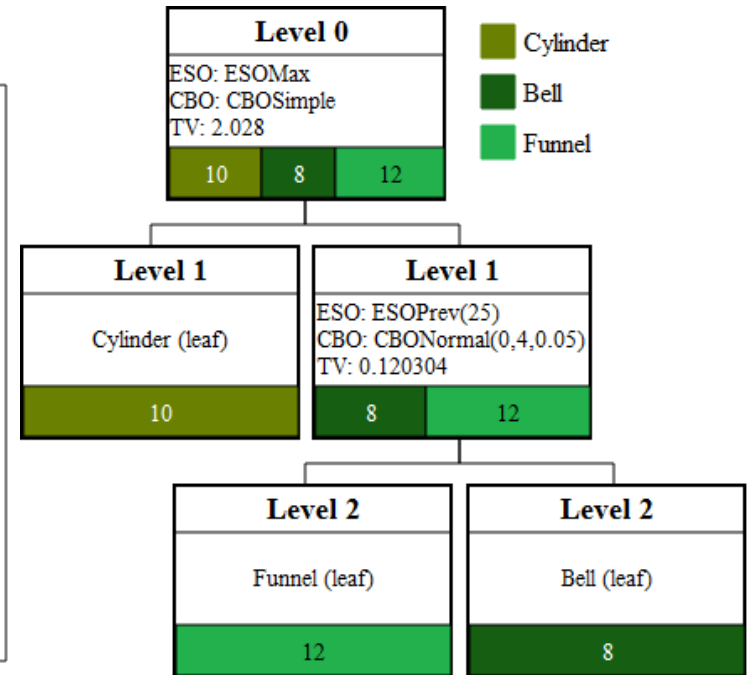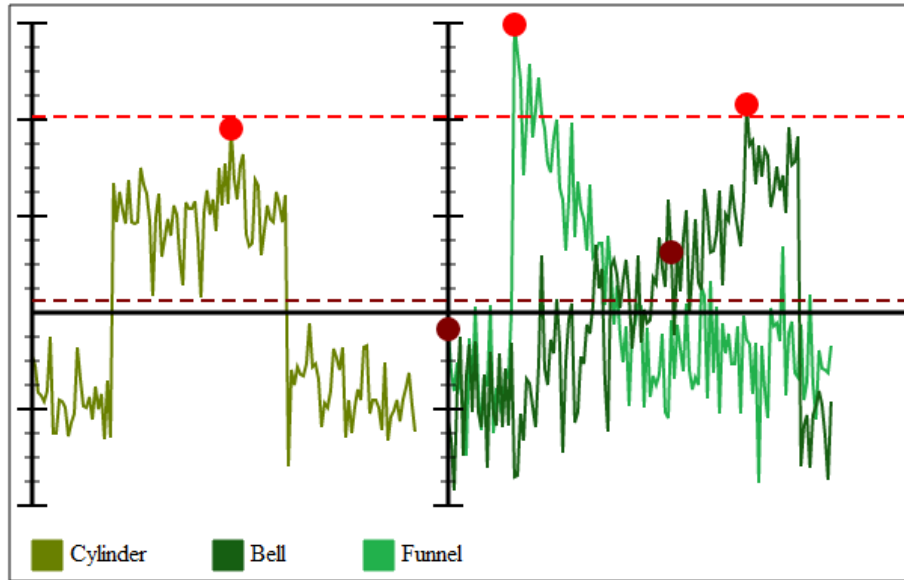
LEARNING (3/3)

- Best split selected
- Cursor is moved by ESO
  - Reachable positions will change

- Train set split
- Child nodes continue the learning process
- Stop: homogenous nodes

Root node

Cursor start: 0

ESO: Go to global max
CBO: Value at time
Threshold: 1,02775

Splitting the set of samples

Left child

Cursor start: 1c

Right child

Cursor start: 1c
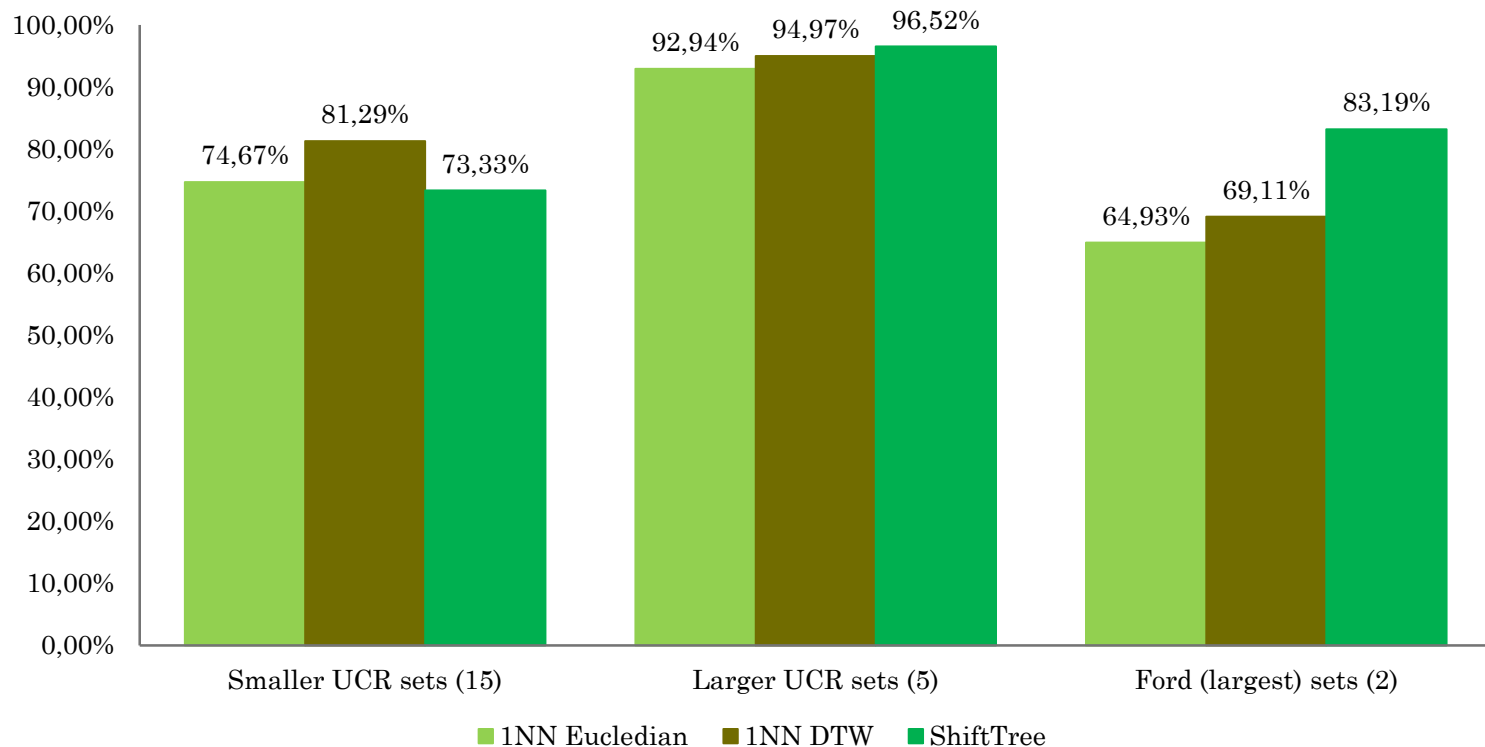
INTERPRETABILITY

- Depends on the operators
- Helps undepstanding the data
- E.g.: CBF dataset
  - Z-normalized, 3 class: Cylinder, Bell, Funnel
  - Distinguish cylinder from bell and funnel by global maxima
    - The data is z-normalized (standardized)
  - Distinguish bell from funnel by stepping back 25 steps + noise filtering through weighted average
    - On which side is the peak

PERFORMANCE OF THE ALGORITHM

- Databases
  - UCR: mostly smaller data sets (20)
  - Ford: larger data sets (2)
- No optimizations in the experiments
- Performance
  - Better on larger data sets (model-based)

Advantages and drawbacks

- Advantages
  - Advantages of being model based (fast labeling, etc)
  - Interpretable (depens on operators)
  - Can be used independently of application domain
  - Expert's knowledge can be built in through operators
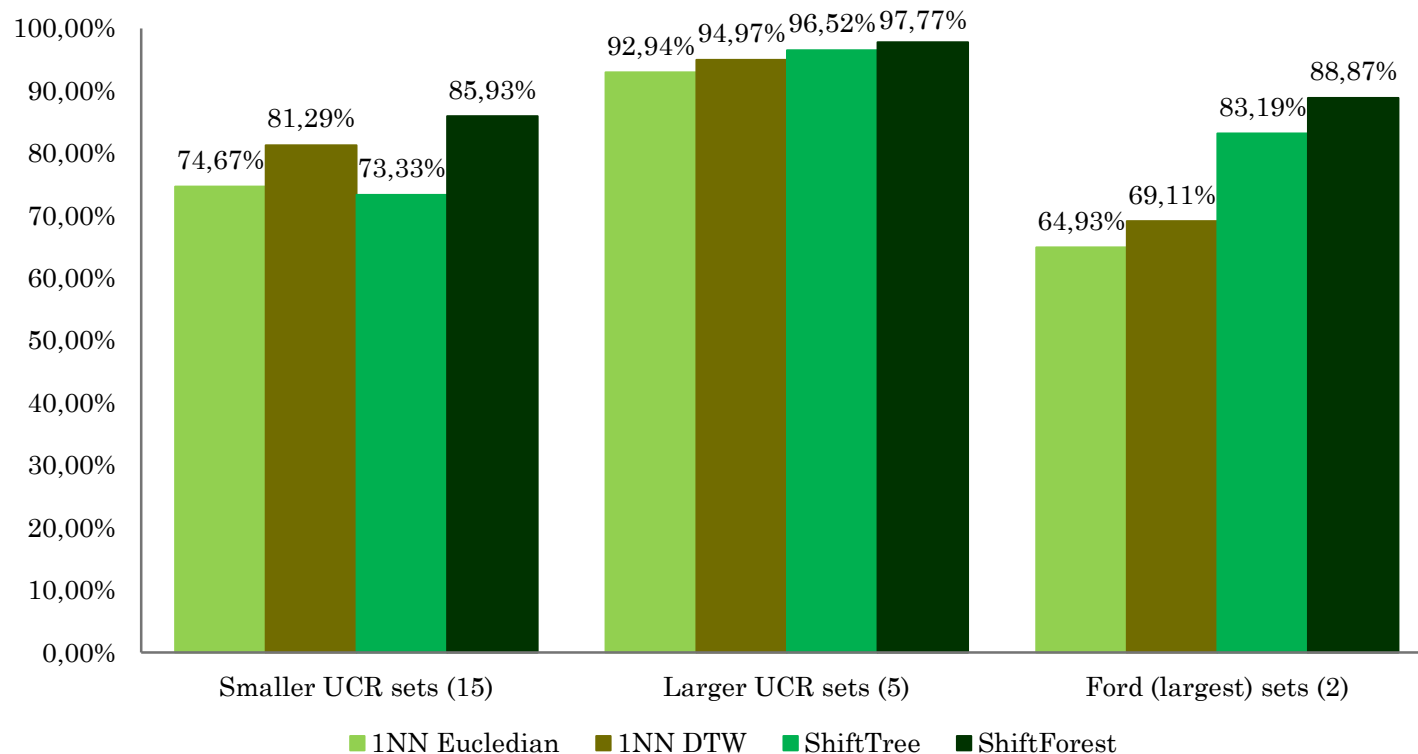
- Disadvantages
  - Optimization of the operators is not trivial
  - Learning may take a long time when too many operators used
  - Disadvantages of the model based methods (larger training set required, etc)

SHIFTFOREST: COMBINING MULTIPLE MODELS

- Combining models enhances accuracy
- Boosting
  - Some issues on smaller datasets

- The „XV" method
  - Splitting training set (random splits)
  - Building model on the first part, evaluating on the second
 - Model weight: the measured accuracy

SHIFTFOREST RESULTS

- Accuracy increases
- Interpretability is often lost
  - May the intersection of the trees can be interpreted

Chart legend: 1NN Eucledian, 1NN DTW, ShiftTree, ShiftForest

	1NN Eucledian	1NN DTW	ShiftTree	ShiftForest
Smaller UCR sets (15)	74,67%	81,29%	73,33%	85,93%
Larger UCR sets (5)	92,94%	94,97%	96,52%	97,77%
Ford (largest) sets (2)	64,93%	69,11%	83,19%	88,87%

Conclusion

- Novel model-based algorithm
- Uses cursors and operators
  - Cursor movement
  - Attribute computation
- Sequence of simple operators as the model
  - Decision tree
- Base of a whole algorithm family

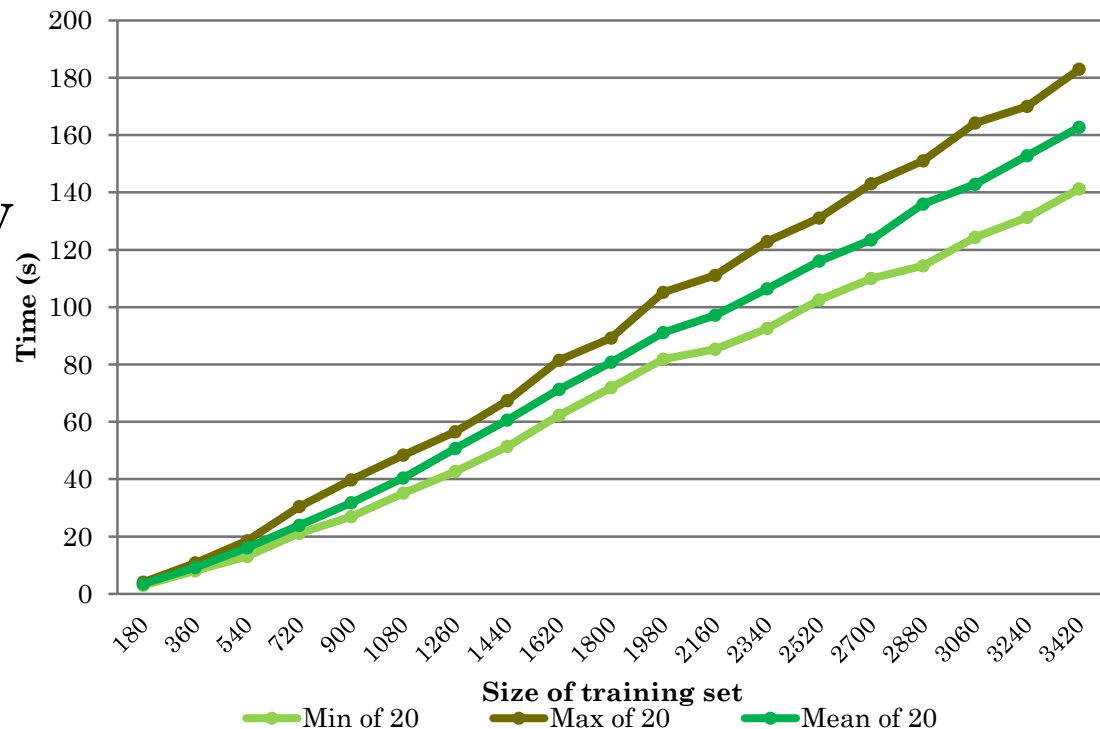# THANK YOU FOR YOUR ATTENTION!
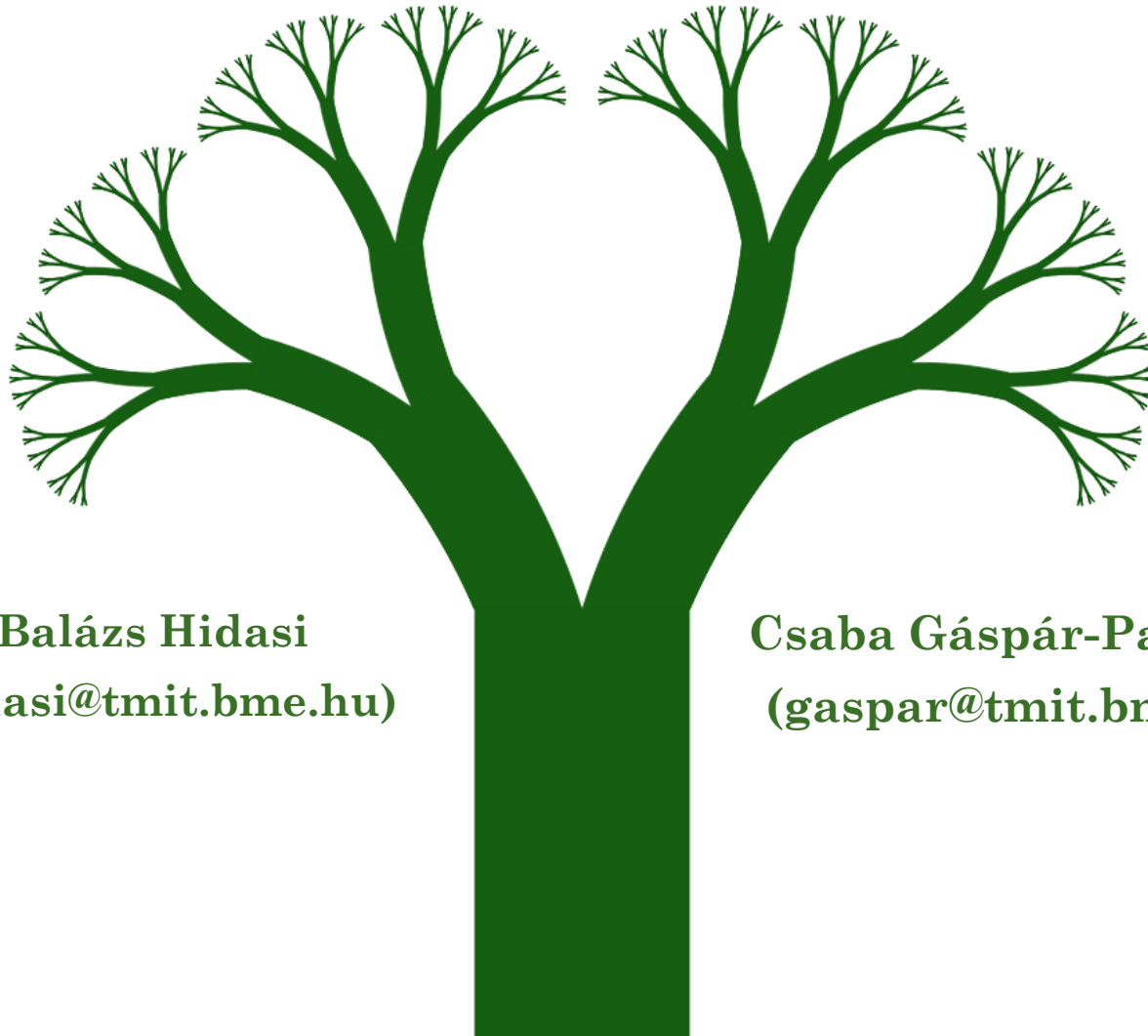
Balázs Hidasi

(hidasi@tmit.bme.hu)

Csaba Gáspár-Papanek

(gaspar@tmit.bme.hu)

TIME OF MODEL BUILDING, SCALABILITY

- O(N*log(N)) per node (ordering)

- Time of model building
  - Depends on tree structure
  - Acceptable

- Scales linearly
  - Experiments
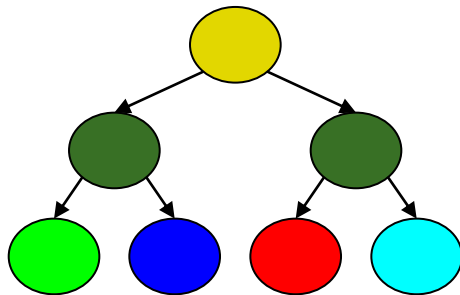
LEARNING THE MODEL

Defined operators		
ESO	ESONextMax	To the next local maxima
	ESONext100	Step forward 100
	ESOMax	To the global maxima
CBO	CBOSimple	Value at cursor position

ShiftTree

Best split in node (so far)	
ESO:	-
CBO:	-
Threshold:	-

EyeShifter – Cursor position
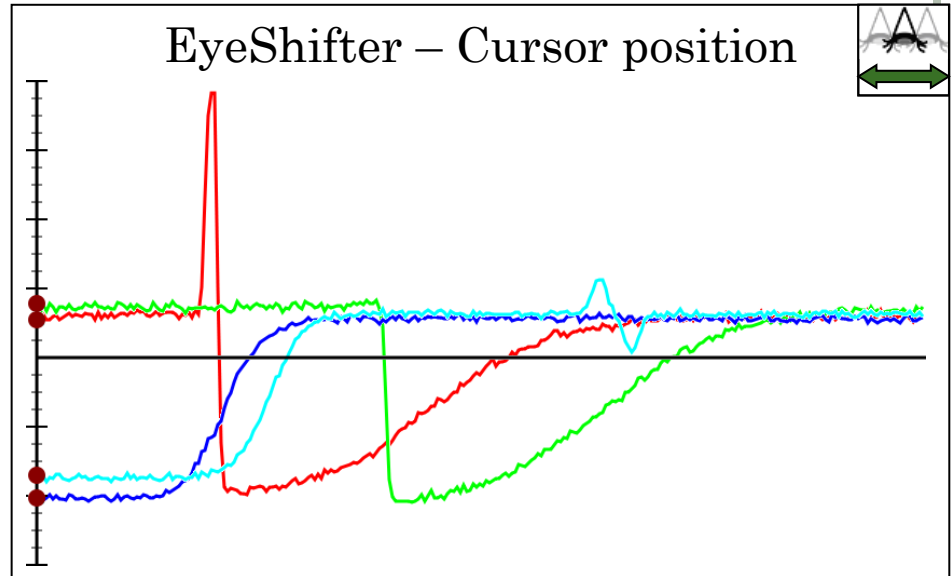
ConditionBuilder - Ordering
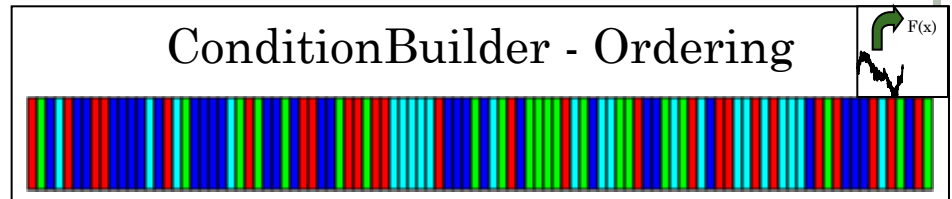
Dynamic attribute selection

Best threshold by current dynamic attribute:	-
Score of the best split by current dyn. attribute	Inf
Best score in the node (so far)	Inf
Best ordering:	