On the Viability of Decision Trees for Learning Models of Systems

Swantje Plambeck, Lutz Schammer,
and Görschwin Fey

Hamburg University of Technology

20.01.2022

Outline

1. Motivation
2. Basics (Automata & Decision Trees)
3. Representing Mealy Machines in Decision Trees
4. Limitations for Exact Learning
5. Practical Assessment
6. Experimental Results
7. Conclusion

Motivation

- Abstract models are useful for many applications like Monitoring, Testing, Simulation, etc.

Learning exact models from limited resources
(specifically bounded history)
is known to be **impossible** [7]

We **prove** this
(using Decision Trees)

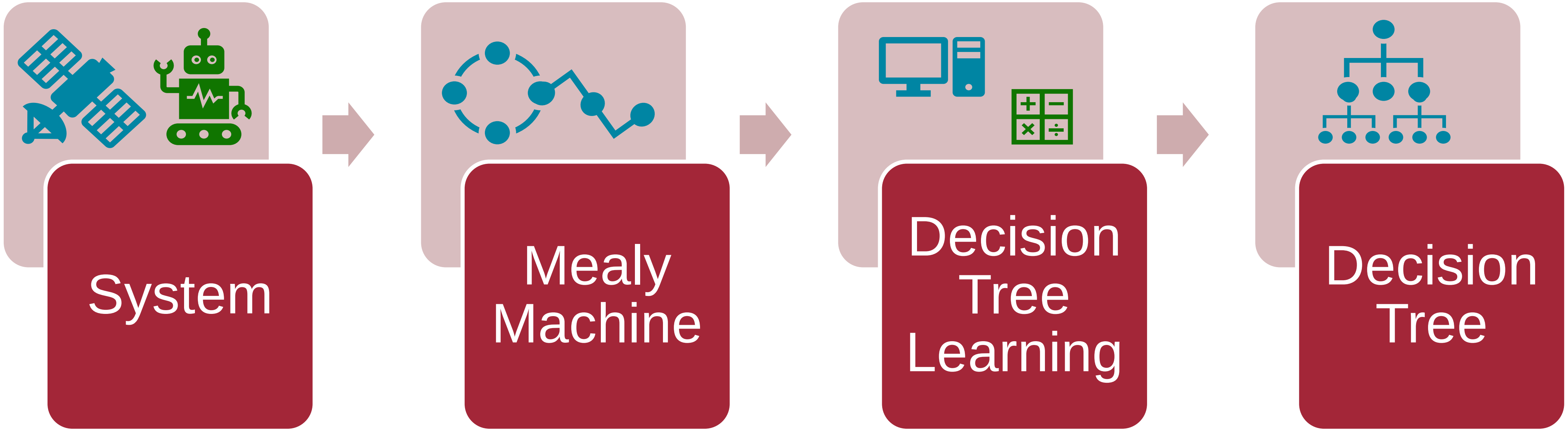
Using limited history is often necessary or required
(and offers efficiency)

We demonstrate that this is possible for **practical** cases

Motivation

- Goal:

  Find a decision tree representation of systems that can be modelled as a Mealy machine.



Motivation

- Goal:

  Find a decision tree representation of systems that can be modelled as a Mealy machine.

- Why do we do this?
  - Decision trees and decision tree learning have been applied in many scenarios already
  - Decision tree learning is efficient
  - Decision trees interpretable and assumingly more flexible

- We determine whether a perfect reconstruction as a decision tree is possible
- If no perfect reconstruction is possible: what do we lose?

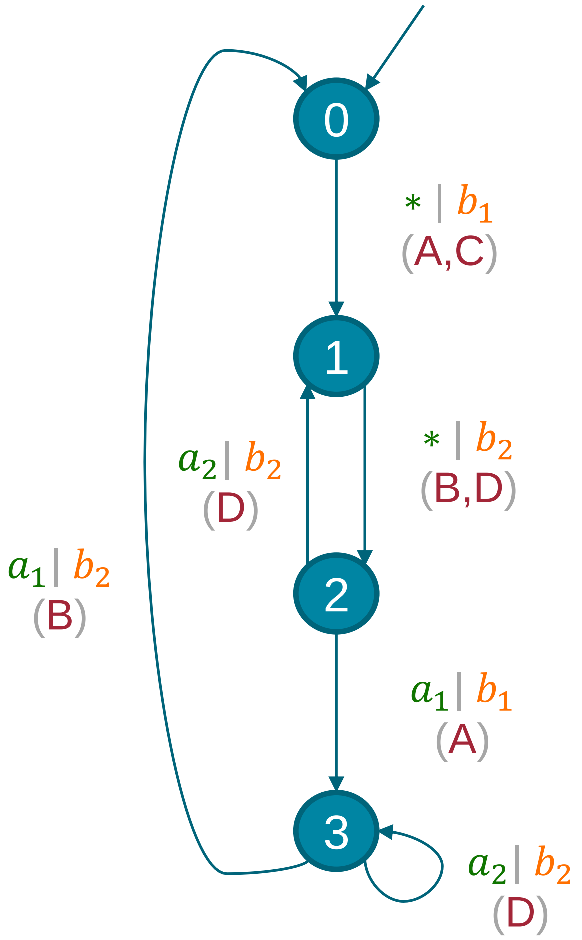
Basics – Automata and Decision Trees

Mealy machines

- Well-established exact approach: automata learning [1,8]
  - often infeasible
  - limited potential for approximations
- We consider observation of bounded history
  - Knowledge about initial state not needed

Decision trees [3,4]

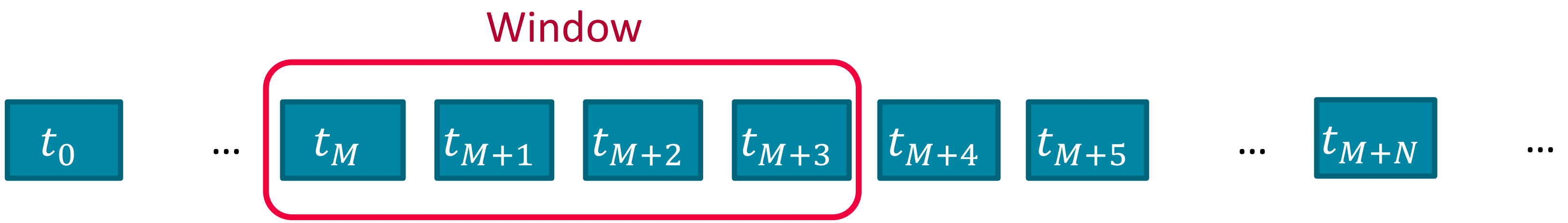
- Decision trees represent a classification function
- Decision tree learning (DTL) requires a set of labelled feature vectors



Representing Mealy Machines in Decision Trees

Mealy Machine	Decision Tree
displays temporal behavior of a system	needs a fixed number of features

Use bounded history of N time steps as features for DTL



Representing Mealy Machines in Decision Trees

Mealy Machine	Decision Tree
displays temporal behavior of a system	needs a fixed number of features

Use bounded history of N time steps as features for DTL

Determines next state and output on given input	classifies feature vectors

Class label is next output

Sequence of length $N+1$	Feature vector of length $N+1$ + class label

i_M o_M | i_{M+1} o_{M+1} | i_{M+2} o_{M+2} | i_{M+N} o_{M+N} $\Rightarrow$ $\boxed{i_M\ o_M\ i_{M+1}\ o_{M+1}\ i_{M+2}\ o_{M+2}\ i_{M+N}}$ $\boxed{o_{M+N}}$

Representing Mealy Machines in Decision Trees

Mealy Machine	Decision Tree

Is this an exact mapping?

If not, why?

In which cases could it be exact?

Sequence of length $N + 1$	Feature vector of length $N + 1$ + class label
i_M o_M i_{M+1} o_{M+1} i_{M+2} o_{M+2} i_{M+N} o_{M+N}	$\boxed{i_M \; o_M \; i_{M+1} \; o_{M+1} \; i_{M+2} \; o_{M+2} \; i_{M+N}}$ $\boxed{o_{M+N}}$

Representing Mealy Machines in Decision Trees

Mealy Machine	Decision Tree
has language L [7]	classifies feature vectors

Classes are output or ‚*not in language* L'

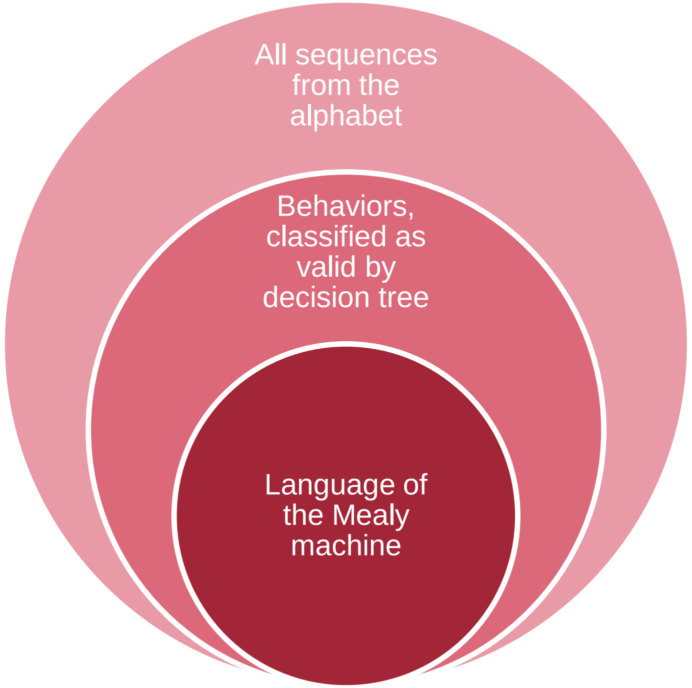
Mapping:

1. Consider all sequences of length $N + 1$ from the alphabet of the Mealy machine
2. Create a feature vector for each sequence
3. Label those sequences, which exist in the Mealy machine with their next output
4. Label those sequences, which are not conform with the Mealy machine‘s language with ‚*invalid*‘

Limitations for Exact Learning

Theorem:

When learning a decision tree from all observations, then the learned representation overapproximates the original Mealy machine.

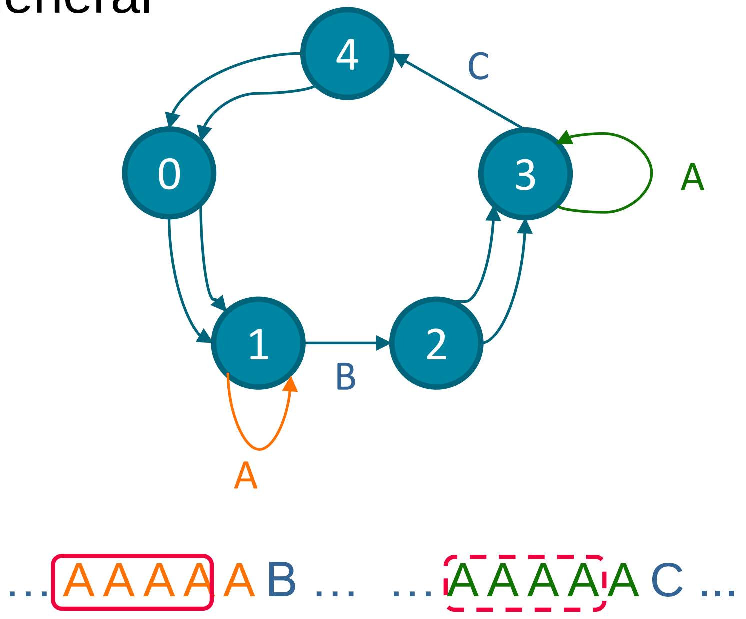


Limitations for Exact Learning

Theorem:

When learning a decision tree from all observations, then the learned representation overapproximates the original Mealy machine.

- Full equivalence of the two representations is not possible in general
- Counterexample:
  - Bounded history eventually not sufficient to identify the system’s state definitely
  - Orientation in the automaton is lost due to limited history



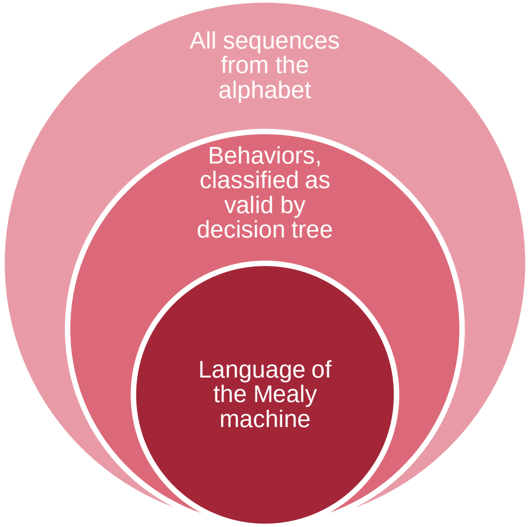
Limitations for Exact Learning

Theorem:

When learning a decision tree from all observations, then the learned representation overapproximates the original Mealy machine.

- Argumentation is independent of decision trees
  - ➢ No guarantee for exact abstraction with limited history in general cases
- Limited history is successfully used in many practical applications
  - ➢ Why is this still possible in practice?

Practical Assessment



Approaches to overcome the described problem:

1. Limitations to the set of sequences on which equivalence holds
   - Only sequences of length N, starting in the initial state
   - Only loop-free sequences
   - Only sequences where each loop is taken at most K times where K is a finite integer

2. Restrictions to the system under learning
   - Any state is uniquely identified by an observation of bounded history
     - ➢ no subsequence of length N precedes two or more different states in the system
     - ➢ different modes of operation where looping behaviour occurs have unique input-output sequences

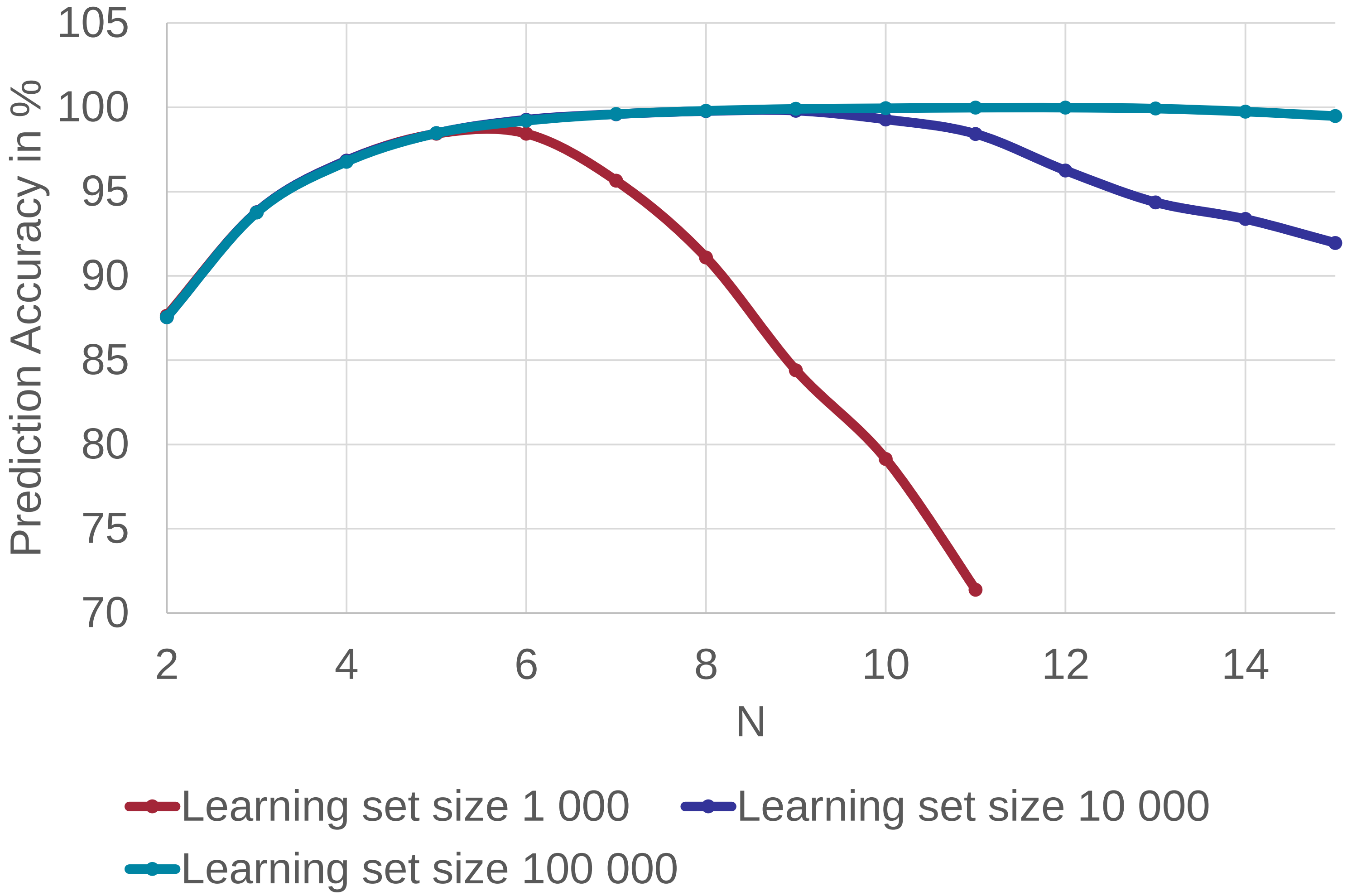
➢ Despite theoretical limitations, decision tree learning decision tree learning is a useful approach achieving good results in several applications

Experimental Results – Language Identification [5,6]

Example	Window Width N	$\frac{\#\text{ learning sequences}}{\#\text{ all sequences}}$	True Positives (%)	False Negatives (%)	True Negatives (%)	False Positives (%)
Simple Example	3	1.00	100	0	93	7
Simple Example	5	1.00	100	0	99	1
Critical Example	1	1.00	100	0	92	8
Critical Example	5	1.00	100	0	≈100	<0
Y86 CPU [2]	3	<0.001	100	0	97	3
Y86 CPU [2]	5	<0.001	100	0	100	0
Coffee Machine [1]	5	0.013	98	2	99	1
Generic (20 states, 14 symbols)	20	<0.001	97	3	100	0
Generic (100 states, 4 symbols)	5	0.898	100	0	0	100
Generic (100 states, 4 symbols)	10	0.069	100	0	100	0

- Many sequences used for learning → no false negatives
- For less observations → false negatives occur
- Increasing window width → better classification (larger number of true negatives)

Experimental Results – Prediction Accuracy [5,6]



- The prediction accuracy increases when more history is considered
- For very large N the prediction accuracy decreases again, because the number of considered sequences with respect to all possible sequences decreases
- This happens earlier, the smaller the learning set size

Conclusion & Outlook

- Decision trees successfully model practical systems
- Restrictions to language or systems ensure exact models with decision trees
- Theoretical results also hold for other learners with bounded history
- Decision trees can serve to predict an upcoming output

- Decision trees allow for learning of approximate models
- Specification of errors from bounded history or limited observations with decision trees

References

[1] M. Merten, “Active automata learning for real life applications,” Ph.D. dissertation, 01 2013.
[2] R. E. Bryant, Computer systems: a programmer’s perspective, D. R. O’Hallaron, Ed. Upper Saddle River, NJ u.a.: Prentice-Hall, 2003.
[3] L. Breiman, J. Friedman, R. Olshen, and C. Stone, Classification And Regression Trees. Wadsworth: Belmont, 1984.
[4] S. Shalev-Shwartz and S. Ben-David, Understanding Machine Learning: From Theory to Algorithms. Cambridge University Press, 2014.
[5] M. Isberner, F. Howar, and B. Steffen, “The open-source learnlib,” in Computer Aided Verification, D. Kroening and C. S. Păsăreanu, Eds. Cham: Springer International Publishing, 2015, pp. 487–495.
[6] MATLAB, (R2020b). Natick, Massachusetts: The MathWorks Inc., 2020.
[7] E. M. Gold, “Language identification in the limit,” Information and Control, vol. 10, no. 5, pp. 447 – 474, 1967.
[8] D. Angluin, “Learning regular sets from queries and counterexamples,” Information and Computation, vol. 75, no. 2, pp. 87 – 106, 1987.