Finding small interpretable ML models for partially defined Boolean functions

Stefan Szeider
Joint work with Eduard Eiben, Sebastian Ordyniak, Giacomo Paesani

TECHNISCHE UNIVERSITÄT WIEN
Vienna | Austria

ALGORITHMS AND COMPLEXITY GROUP

Boolean Seminar
Liblice, Sept 2023

Classification Instances

- Let F be a finite set of **features**.
- An (binary) **example** e is a mapping $e : F \to \{0,1\}$
- A **classification instance** is a pair $E = (E^+, E^-)$ of two disjoint sets of examples, the positive and negative examples, respectively.

features

examples

A	B	C	D	out
0	0	0	0	1
0	0	0	1	?
0	0	1	0	1
0	0	1	1	?
0	1	0	0	1
0	1	0	1	?
0	1	1	0	1
0	1	1	1	?
1	0	0	0	0
1	0	0	1	?
1	0	1	0	1
1	0	1	1	1
1	1	0	0	0
1	1	0	1	?
1	1	1	0	0
1	1	1	1	?

aka Partially Defined Boolean Functions

MR2305784 - Renamable interval extensions of partially defined Boolean functions
Čepek, Ondřej; Kronus, David; Kučera, Petr
WSEAS Trans. Math. **6** (2007), no. 4, 559–566.

MR1951045 - Variations on extending partially defined Boolean functions with missing bits
Boros, Endre; Ibaraki, Toshihide; Makino, Kazuhisa
Inform. and Comput. **180** (2003), no. 1, 53–70.

MR1739062 - Partially defined Boolean functions with applications to data analysis: a survey
Ibaraki, Toshihide
Math. Japon. **51** (2000), no. 1, 153–165.

MR1724734 - Inner-core and outer-core functions of partially defined Boolean functions
Makino, Kazuhisa; Ibaraki, Toshihide
Discrete Appl. Math. **96/97** (1999), 443–460.

MR1698946 - Horn extensions of a partially defined Boolean function
Makino, Kazuhisa; Hatanaka, Ken-Ichi; Ibaraki, Toshihide
SIAM J. Comput. **28** (1999), no. 6, 2168–2186.

MR1604212 - Error-free and best-fit extensions of partially defined Boolean functions
Boros, Endre; Ibaraki, Toshihide; Makino, Kazuhisa
Inform. and Comput. **140** (1998), no. 2, 254–283.

MR1444945 - Positive and Horn decomposability of partially defined Boolean functions
Makino, Kazuhisa; Yano, Kojin; Ibaraki, Toshihide
Discrete Appl. Math. **74** (1997), no. 3, 251–274.

MR1355573 - Decomposability of partially defined Boolean functions
Boros, Endre; Gurvich, Vladimir; Hammer, Peter L.; Ibaraki, Toshihide; Kogan, Alexander
Discrete Appl. Math. **62** (1995), no. 1-3, 51–75.

MR0996267 - Cause-effect relationships and partially defined Boolean functions
Crama, Yves; Hammer, Peter L.; Ibaraki, Toshihide
Ann. Oper. Res. **16** (1988), no. 1-4, 299–325.

MR0550788 - A method for minimizing partially defined Boolean functions
Móricz, F.; Varga, A.; Ecsedi-Tóth, P.
Acta Cybernet. **4** (1978), no. 3, 283–290.

MR0373777 - Elimination of the M-hazard in realizations of partially defined Boolean functions
Sapiecha, Krzysztof
Arch. Automat. i Telemech. **19** (1974), 199–205.

MR0274361 - Reduced disjunctive normal forms of partially defined Boolean functions
Kornev, Yu. N.
Kibernetika (Kiev) **1967** (1971), no. 1, 10–15.
Cybernetics **3** (1967), no. 1, 8–12.

MR0307828 - Realization of partially-defined Boolean functions by expansions in orthogonal series
Karpovskiĭ, M. G.; Moskalev, È. S.
Avtomat. i Telemeh.(1970), no. 8, 88–99.
Automat. Remote Control(1970), no. 8, 1278–1287.

… and many more

Support Sets

- A set $S \subseteq F$ is a **support set** for $E = (E^+, E^-)$ if for each $p \in E^+$ and $n \in E^-$ there is a feature $f \in S$ such that $f(p) \neq f(n)$.
- Finding a smallest support set is NP-hard [Ibaraki, Crama, Hammer, 2011].

Maximum Difference

- The **difference** $\delta(e, e')$ of two examples e, e' is the number of features $f \in F$ such that $f(e) \neq f(e')$.
- The **maximum difference** $\delta(E)$ of a classification instance $E = (E^+, E^-)$ is the maximum $\delta(n, p)$ over all pairs $p \in E^+$ and $n \in E^-$.
- In practice, the maximum difference is often quite small.

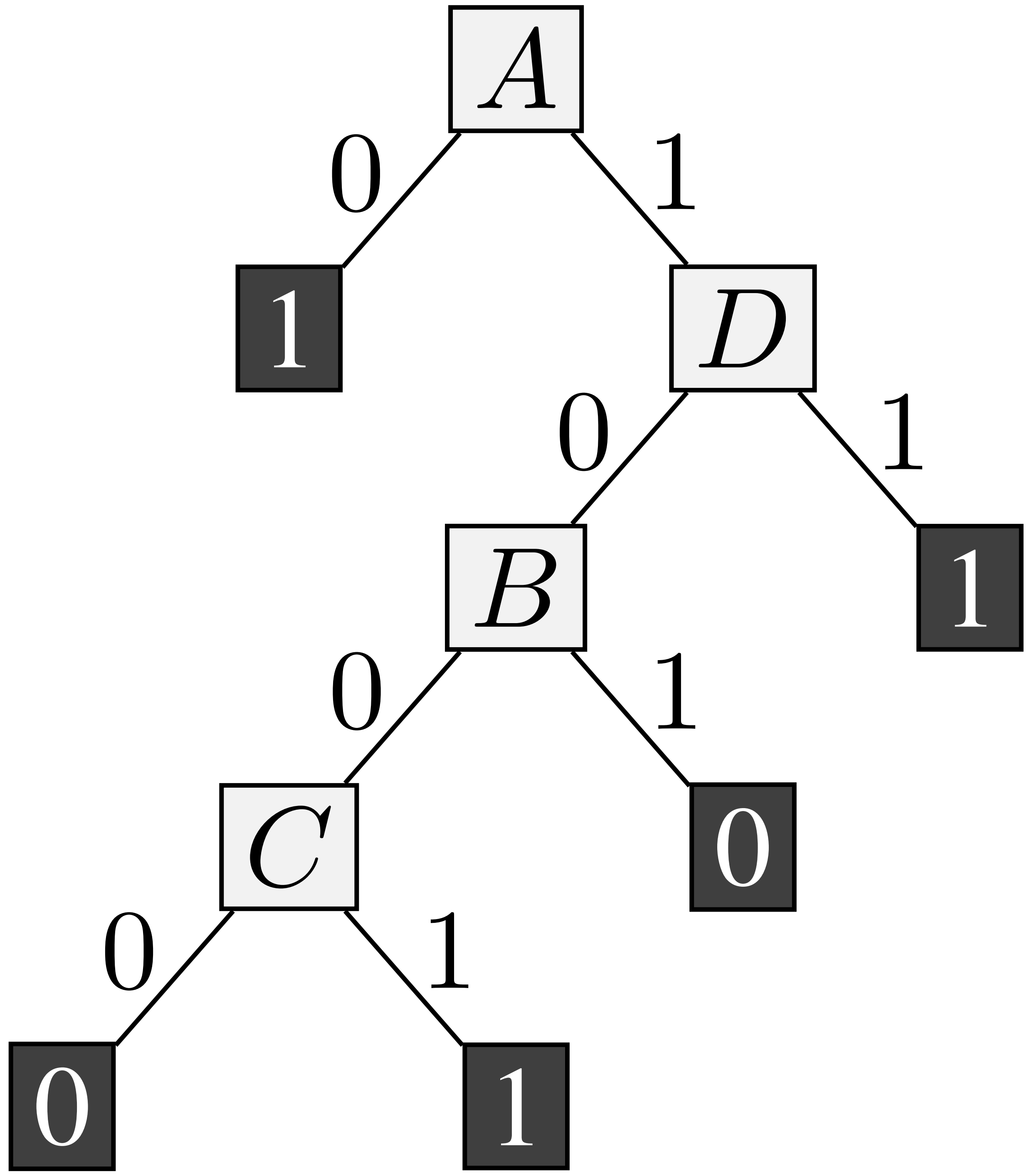
instance	examples	features	max difference
append-tis	106	531	14
australian	690	1164	24
backache	180	476	31
car	1728	22	12
cancer	683	90	18
colic	368	416	43
cleve	303	396	23
haberman	300	93	6
heart	270	382	23
hepatitis	155	362	34
house-	435	17	16
hungarian	294	331	24
new-tyroid	215	335	10
promoters	106	335	106
shuttle	14500	692	18
spect	250	23	22

Narodytska et al. 2018, UCI ML repo

Decision Trees

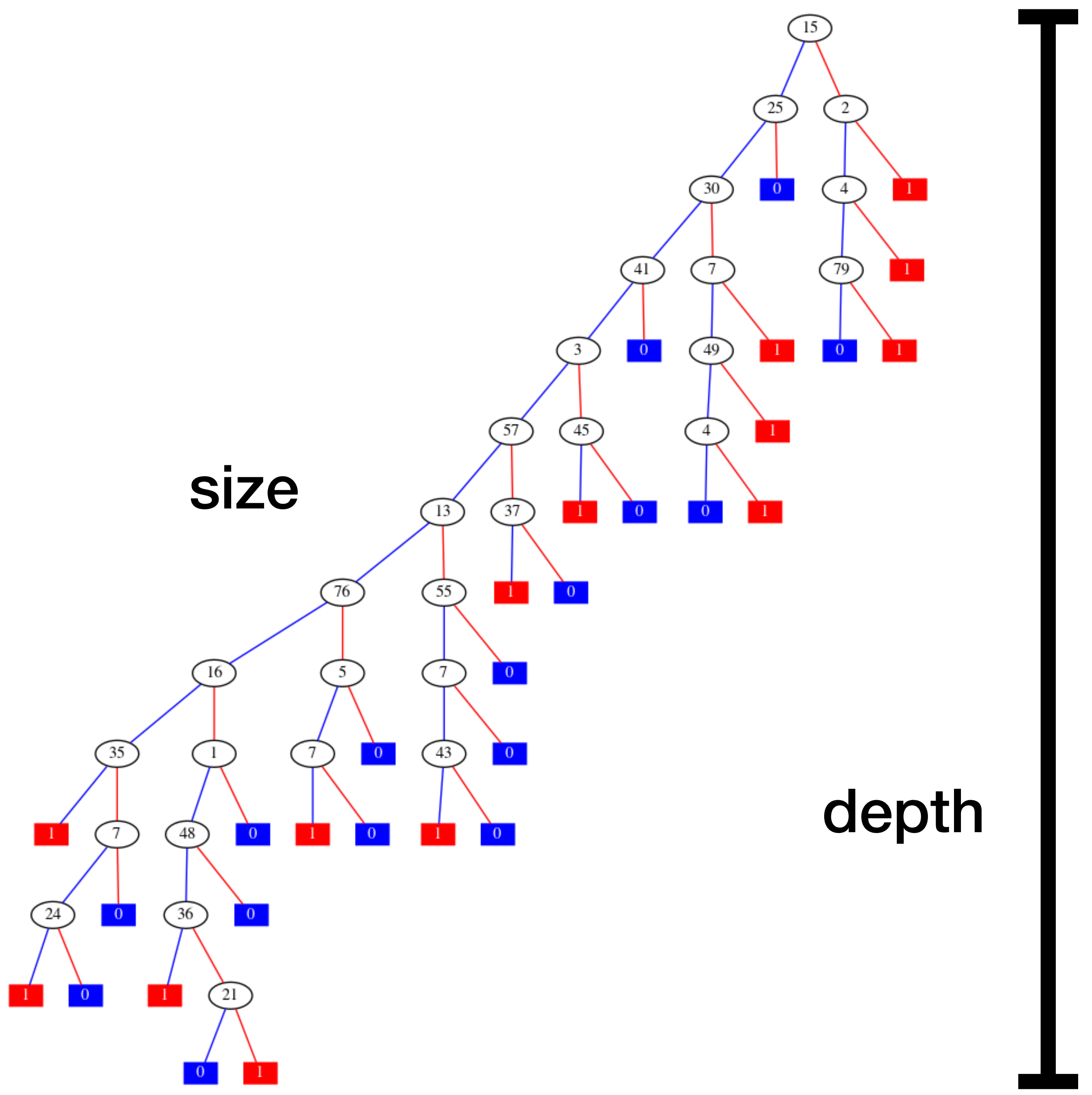
- A DT is a binary rooted tree whose nodes are labeled by features.
- Each leaf is labeled 0 or 1.
- Each non non-leaf has a 0-child and a 1-child.
- A DT T **classifies** a CE E if each positive example ends up in a 1-leaf and each negative example ends up in a 0-leaf.
- Notation: $T \models E, T \models e$
- The DT does not need to use all the features, just the features of a support set.

A	B	C	D	out
1	0	1	1	1
1	0	0	0	0
1	1	0	0	0
1	1	1	0	0
1	0	1	0	1
0	0	0	0	1
0	0	1	0	1
0	1	0	0	1
0	1	1	0	1



Interpretable Models

- Subsymbolic models like neural networks are opaque und difficult to understand.
- DTs and other symbolic ML models have gained new interest.
- Still, a large DT is still difficult to understand:
- Hence want to find a DT of **smallest size** (and still correctly classifies the CE).
- In some cases we want to find a DT of **smallest depth,** limiting the maximum number of tests that need to be performed for each example (tests can be expensive or risky).



Computational Problems

Min-DT-Size

- Instance: a classification instance $E = (E^+, E^-)$, and $s > 0$
- Question: is there a DT T of size $\leq s$ such that $T \models E$? (size=number of non-leaf nodes)

Min-DT-Depth

- Instance: a classification instance $E = (E^+, E^-)$, and $d > 0$
- Question: is there a DT T of depth $\leq d$ such that $T \models E$? (depth=length of longest root-leaf path)

- Both problems are NP-hard [Hyafil and Rivest 1976].
- Many papers have been published over the last couple of years that uitilize SAT-solvers to find smallest/lowest-depth DTs.
- Parameterized Complexity?

Parameterized Complexity of Min-DT-size and Min-DT-height

[Ordyniak-Sz AAAI 2021]

parameters		complexity
solution size	**max difference**	
		para-NP-hard
✓		W[2]-hard, in XP
	✓	para-NP-hard
✓	✓	FPT

- Hardness results: reduction from Hitting Set

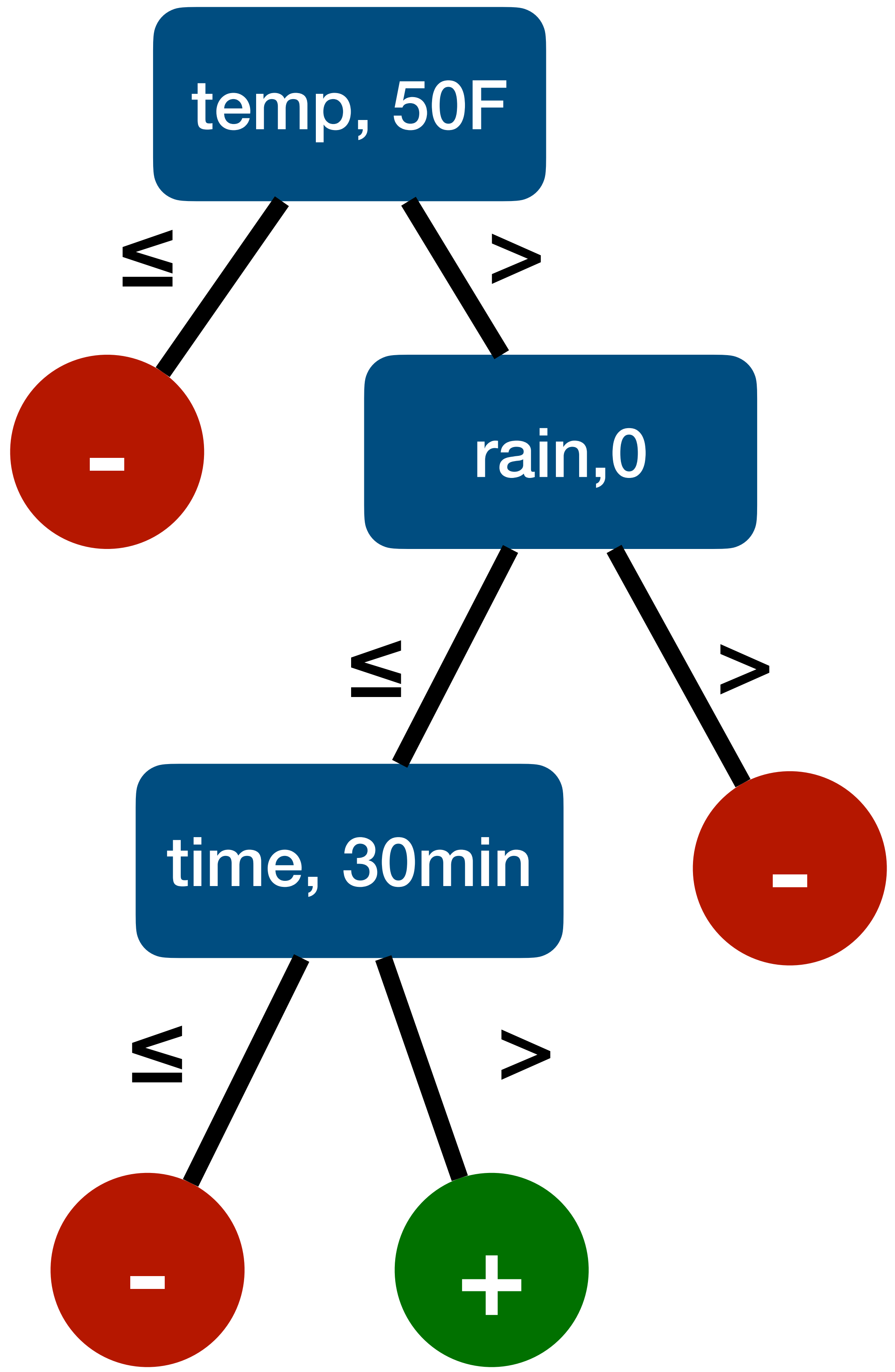
Tractability Results

- **Theorem:** if a support set S and an integer s are given, then we can find in time $2^{s^2}\text{poly}(n)$ a DT using exactly the features in S of size $\leq s$ or decide that it doesn't exist.

- As we can enumerate all S of size $\leq s$ in time n^s, this gives us the XP-results.
- We can enumerate all **minimal** S of size $\leq s$ in time $\delta^s \cdot n$
- The features used in a smallest DT is not necessarily minimal.
- But we can enumerate in FPT time all extensions of a minimal support set that result in a smaller DT. Repeat the extension process. This gives the FPT results.

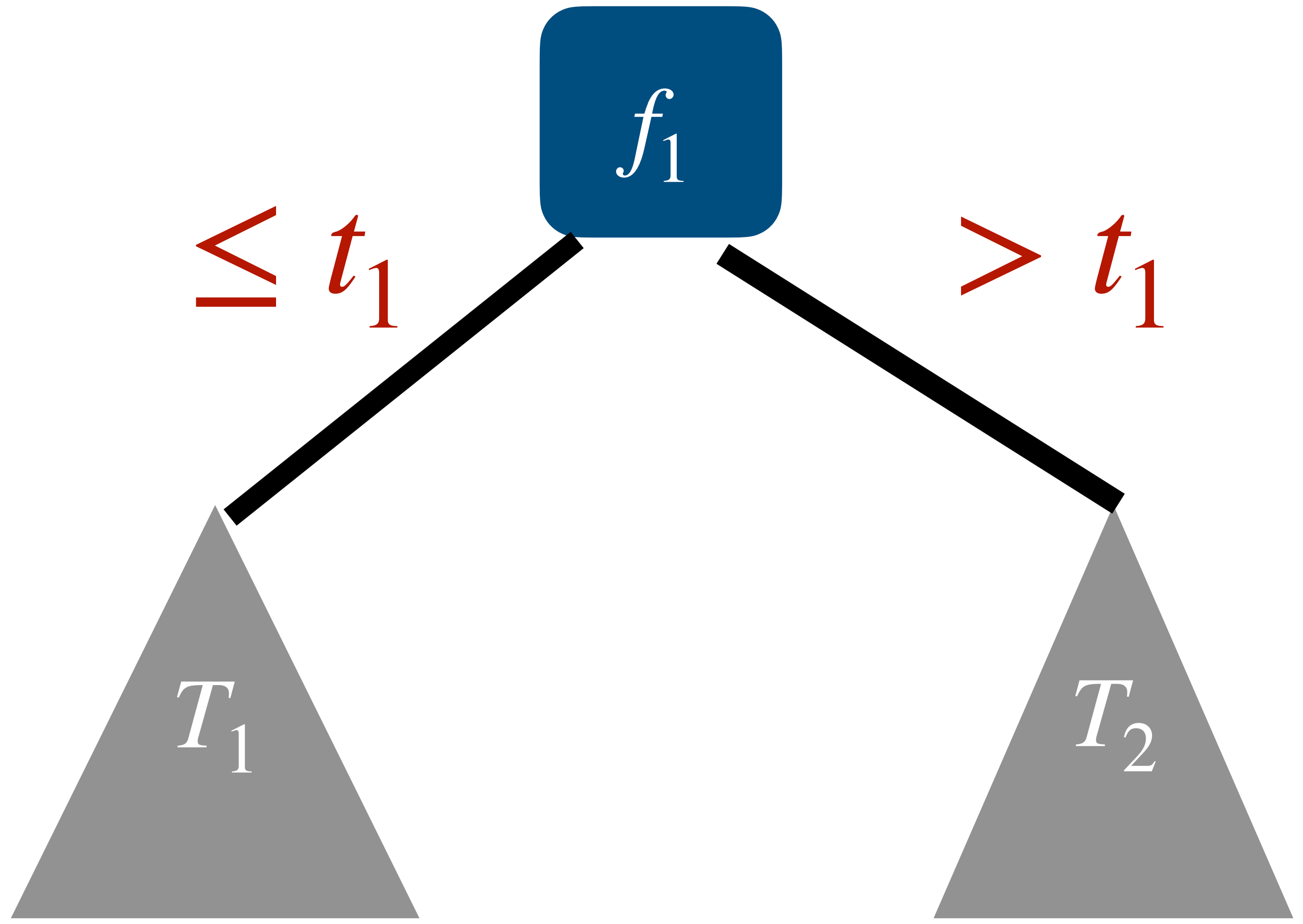
Above the Boolean Case

- Consider features that range over an ordered domain of values.
- At each inner node of the DT we branch on a threshold value.
- The hardness results carry over.
- How about the FPT results?



Threshold Selection

- **Idea:** start with a small value for threshold t_i and gradually increase it
- This way, more and more samples move from T_2 to T_1.
- The size of an optimal T_1 increases. Stop just before it gets too large.
- This idea can be improved by using **binary search** on the threshold value.
- Gives FPT with parameters
    - solution size
    - maximum difference
    - ~~maximum domain size.~~



[Eiben, Ordyniak, Paesani, Sz. IJCAI'23] with a refined notion of branching sets

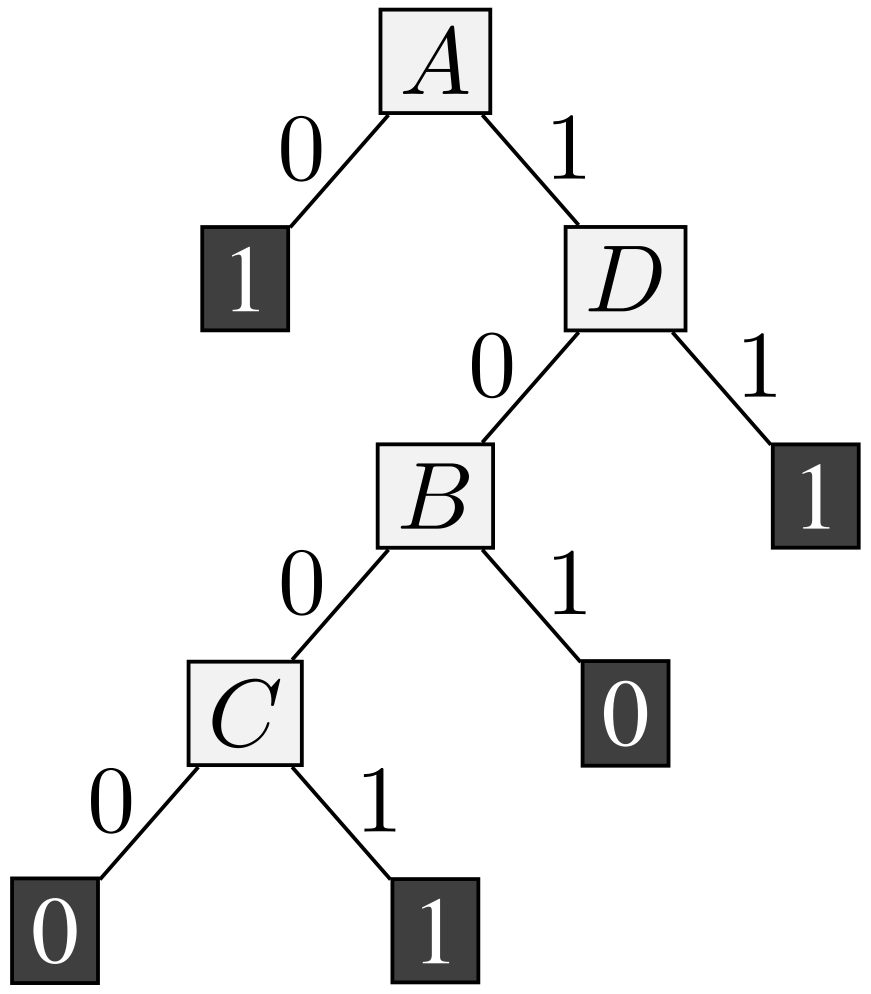
Other Symbolic Models

is min-Model-size parameterized by (solution size + maximum difference) FPT for all these models?

A	B	C	D	out
1	0	1	1	1
1	0	0	0	0
1	1	0	0	0
1	1	1	0	0
1	0	1	0	1
0	0	0	0	1
0	0	1	0	1
0	1	0	0	1
0	1	1	0	1

CE
classification instance

Size:



DT
decision tree

number of decision nodes

if $A = 0$ then 1
if $A = 1 \land B = 0 \land C = 1$ then 1
default 0

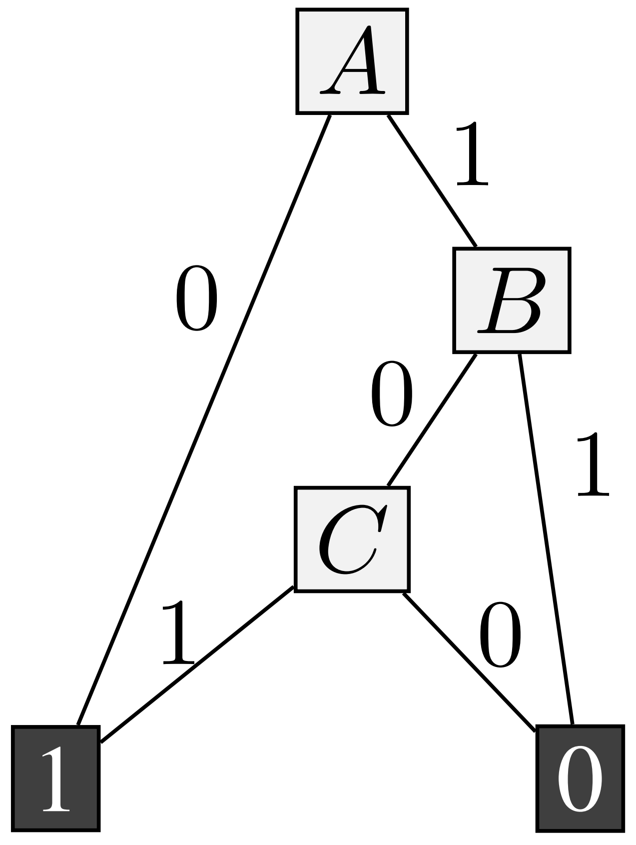
DS
decision Set

number of literals

if $A = 0$ then 1
elseif $B = 1$ then 0
elseif $C = 1$ then 1
elseif true then 0

DL
decision List

number of literals



BDD
binary decision diagram

number of decision nodes

General Results

- **Model type T $\in \{$DT, DS, DL, BDD$\}$**
- **Min-T-Size**
  - Instance: a classification instance $E = (E^+, E^-)$, and $s > 0$
  - Question: is there a model M of type T of size $\leq s$ such that $M \models E$?

- **Theorem:**

  Min-T-Size is FPT for all T $\in \{$DT, DS, DL, BDD$\}$.

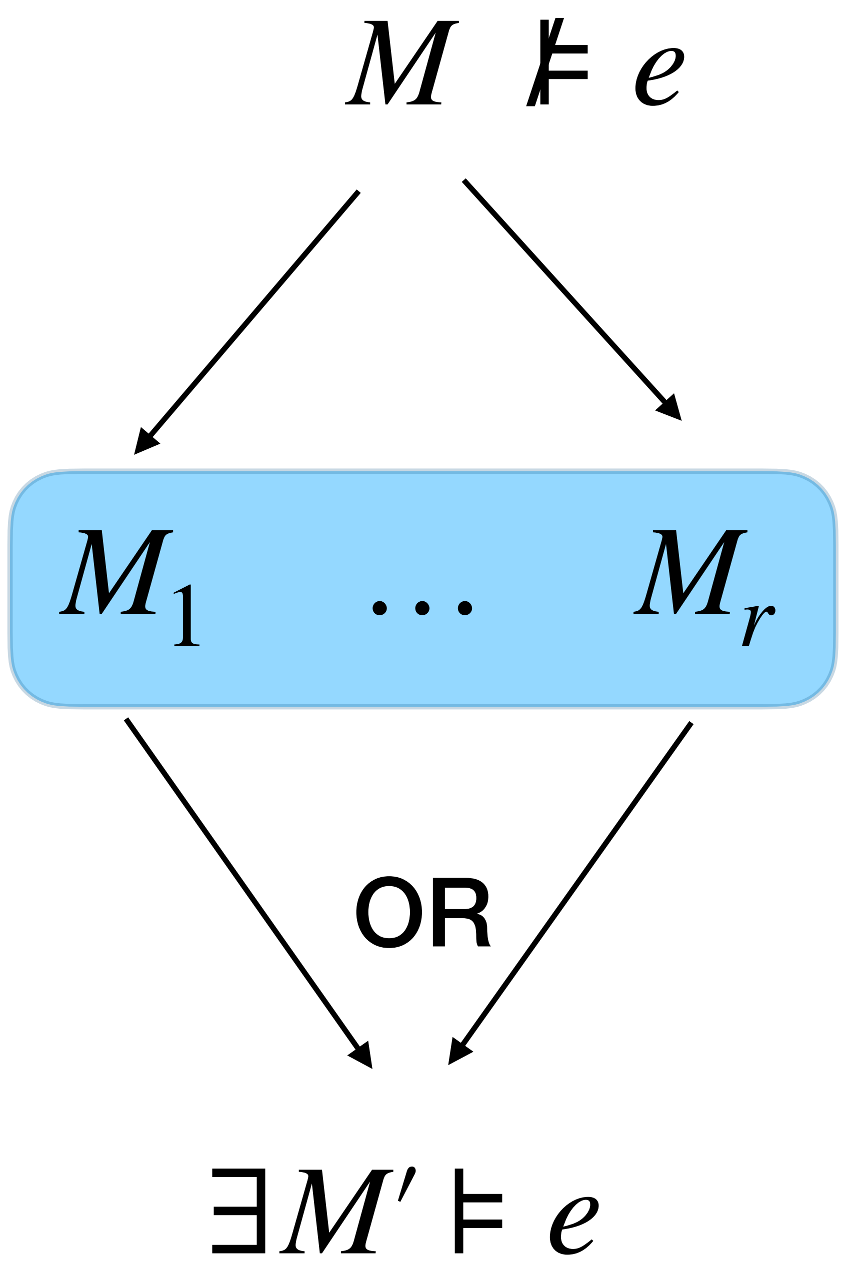
Via meta-theorem based on extendability!

Approach

- Start with empty model.
- As long as some examples are incorrectly classified, branch into certain **extensions** of the current model $(M \longrightarrow M')$.
- Branching is exhaustive.
- Stop if a correct model of size $\leq s$ has been found or return that no such model exists.

- We **annotate** models with examples that guide the selection features one needs to add.
- The algorithmic idea was first used by [Komusiewicz et al. ICML 2023] for DTs.
- We generalize and improve the method.
- We show that it applies to DS, DL, BDD.

Extending a Partial Model



full set of extensions
M_i are annotated models

- **Theorem:** If a full set of extensions for a model of type T can be computed in time $g(n, \delta)$ and has size $f(s, \delta)$ then Min-T-size can be solved in time $f(s, \delta)^s \cdot (g(n, \delta) + n)$.

Results for DS, DL, DT

	$f(s, \delta)$	$g(n, \delta)$	**Min-T-Size**
DS	δ	ns	FPT
DL	$\delta + s + 1$	ns	FPT
DT	$\delta(s + 1)$	ns	FPT

BDDs need extra treatment

Ensembles

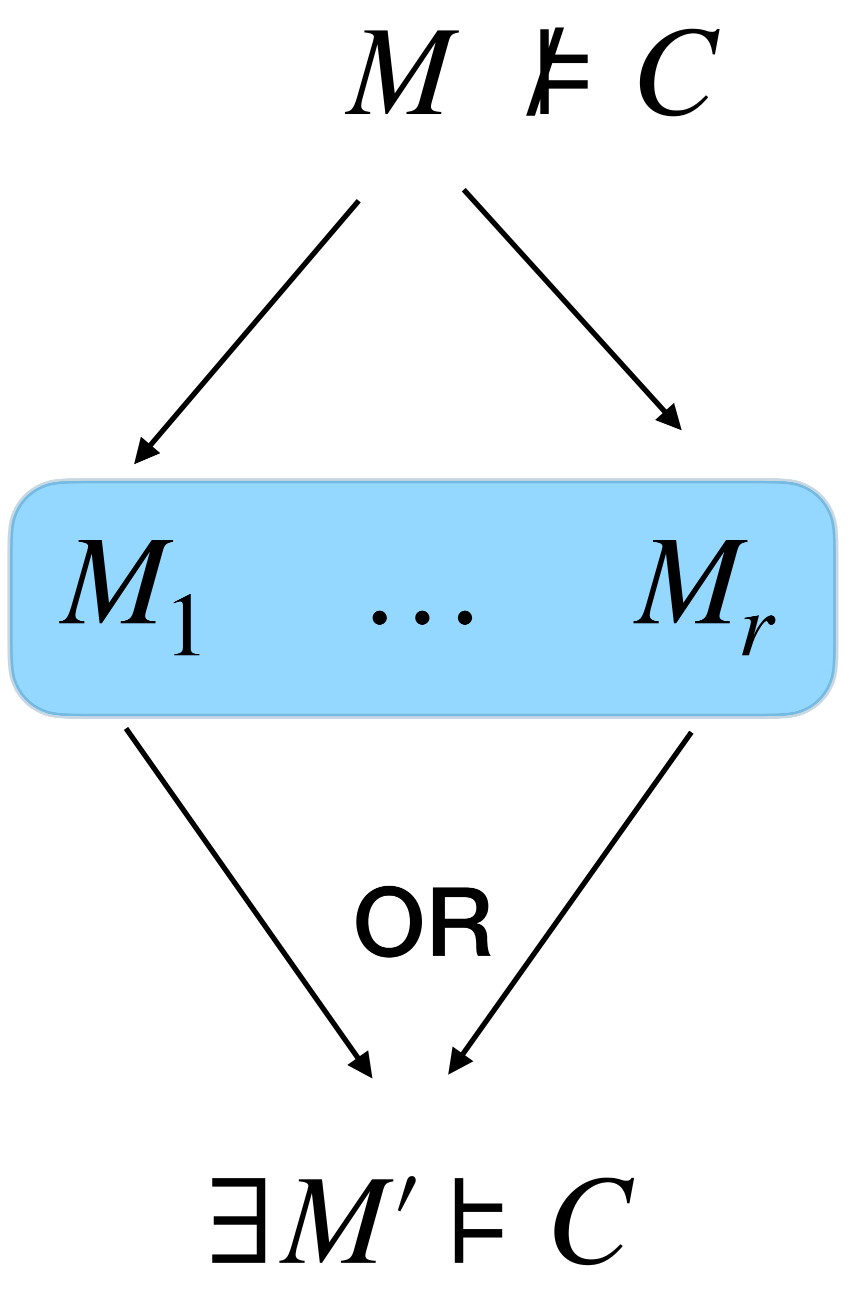
- A **T-ensemble** for a model type T is a set $N = \{M_1, \ldots, M_r\}$ of models of type T.
- A T-ensemble N acts as a single model where N classifies an example according to the majority decision of its element models.
- Let T* denote the model type of T-ensembles
- Size of an ensemble is the sum of sizes of ist element models.
- Ensembles are used in practice, random forests are decision tree ensembles.
- We can extend our meta theorem to ensembles:

- **Theorem:** If a full set of extensions for a model of type T can be computed in time $g(n, \delta)$ and has size $f(s, \delta)$ then Min-T*-size can be solved in time $f(s, \delta)^s \cdot s(g(n, \delta) + n)$.

Results for DS, DL, DT

	$f(s, \delta)$	$g(n, \delta)$	**Min-T-Size**	**Min-T*-Size**
DS	δ	ns	FPT	FPT
DL	$\delta + s + 1$	ns	FPT	FPT
DT	$\delta(s + 1)$	ns	FPT	FPT

Weakly Extending a Partial Model



full set of weak extensions
M_i are models

- **Theorem:** If a full set of weak extensions for a model of type T can be computed in time $g(n, \delta)$ and has size $f(s, \delta)$ then Min-T-size can be solved in time $f(s, \delta)^s \cdot (g(n, \delta) + n)$.

Results for BDDs

	$f(s, \delta)$	$g(n, \delta)$	**Min-T-Size**	**Min-T*-Size**
DS	δ	ns	FPT	FPT
DL	$\delta + s + 1$	ns	FPT	FPT
DT	$\delta(s + 1)$	ns	FPT	FPT
BDD	$\delta 3^{O(s)}$	$2^{O(S)}n^{O(1)}$	FPT	FPT

- Set of weak extensions is much larger, but still FPT-size for BDDs.
- No meta theorem for weak extensions, but we can show that BDD-ensembles have FPT-size weak extensions.

All Results

	$f(s,\delta)$	$g(n,\delta)$	**Min-T-Size**	**Min-T*-Size**
DS	δ	ns	FPT	FPT
DL	$\delta + s + 1$	ns	FPT	FPT
DT	$\delta(s+1)$	ns	FPT	FPT
BDD	$\delta 3^{O(s)}$	$2^{O(S)} n^{O(1)}$	FPT	FPT

- **Theorem:**

  Min-T-Size and Min-T*-Size is FPT for all T $\in \{$DT, DS, DL, BDD$\}$.

Conclusion

- Theory of **partially defined Boolean functions** provides a fruitful foundation for the study of symbolic ML models.
- Provided FPT results for computing smallest models for model types **DT,DS,DL,BDD** and **ensembles** thereof.
- BDD results also apply to special classes (free, ordered, etc.).
- **Meta-results** based on extensions are particularly appealing.
- Results most likely generalize to any **finite ordered domains**, however we expect that domain size needs to be included as a parameter.

Some References

1. Eduard Eiben, Sebastian Ordyniak, Giacomo Paesani, Stefan Szeider: Learning Small Decision Trees with Large Domain. IJCAI 2023: 3184-3192.

2. Sebastian Ordyniak, Stefan Szeider: Parameterized Complexity of Small Decision Tree Learning. AAAI 2021: 6454-6462.

3. Christian Komusiewicz, Pascal Kunz, Frank Sommer, Manuel Sorge: On Computing Optimal Tree Ensembles. ICML 2023: 17364-17374.

4. Sebastian Ordyniak, Giacomo Paesani, Stefan Szeider: A General Theoretical Framework for Learning Smallest Interpretable Models, submitted, 2023.