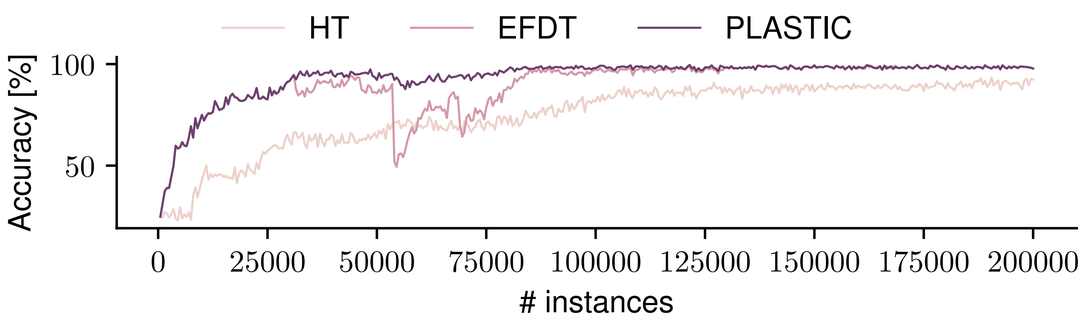
Leveraging Plasticity in Incremental Decision Trees

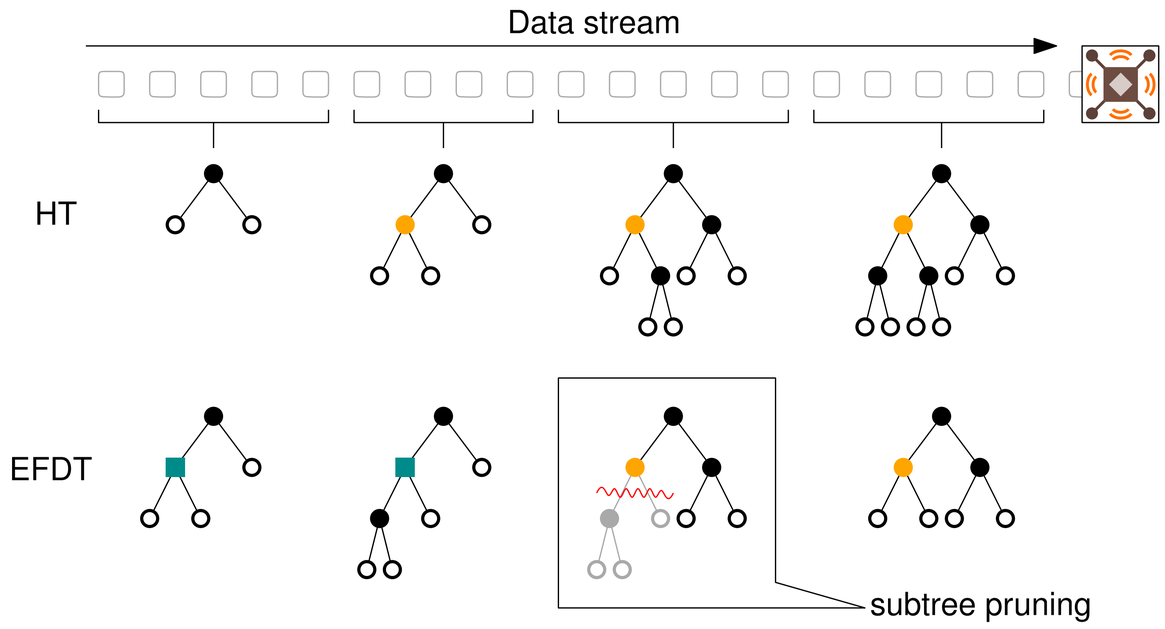
Marco Heyden (marco.heyden@kit.edu), Heitor Murilo Gomes, Edouard Fouché, Bernhard Pfahringer, and Klemens Böhm

Preliminaries

- Hoeffding Trees (HT) [1] and Extremely Fast Decision Trees (EFDT) [2] are commonly used incremental decision trees
- Both rely on statistical testing to decide whether to split a leaf
- EFDT is able to learn from less data than HT
- But EFDT occasionally prunes subtrees if it has found a better split
- **Leads to sudden and unpredictable decreases in accuracy**

Our algorithm PLASTIC restructures the otherwise pruned subtree





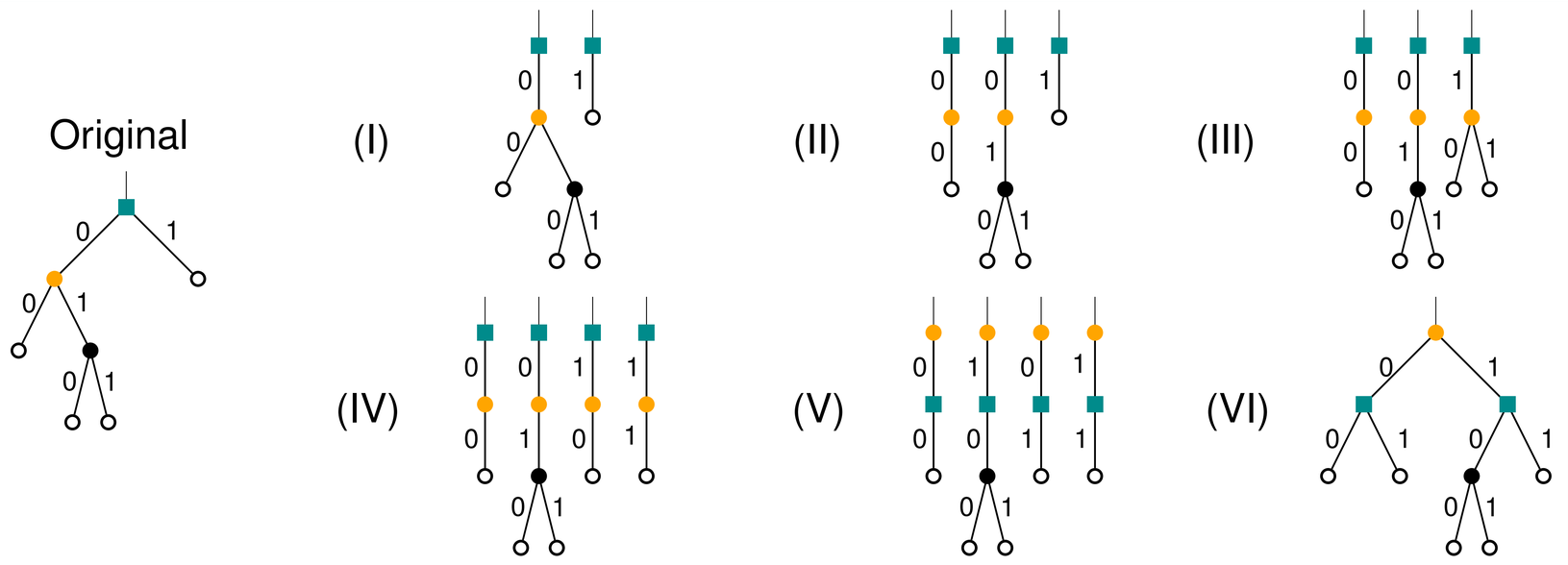
Our algorithm: PLASTIC

Decision tree plasticity

- One can alter the structure of a decision tree without affecting predictions
- Allows restructuring a sub-tree s.th. the root splits at the desired attribute

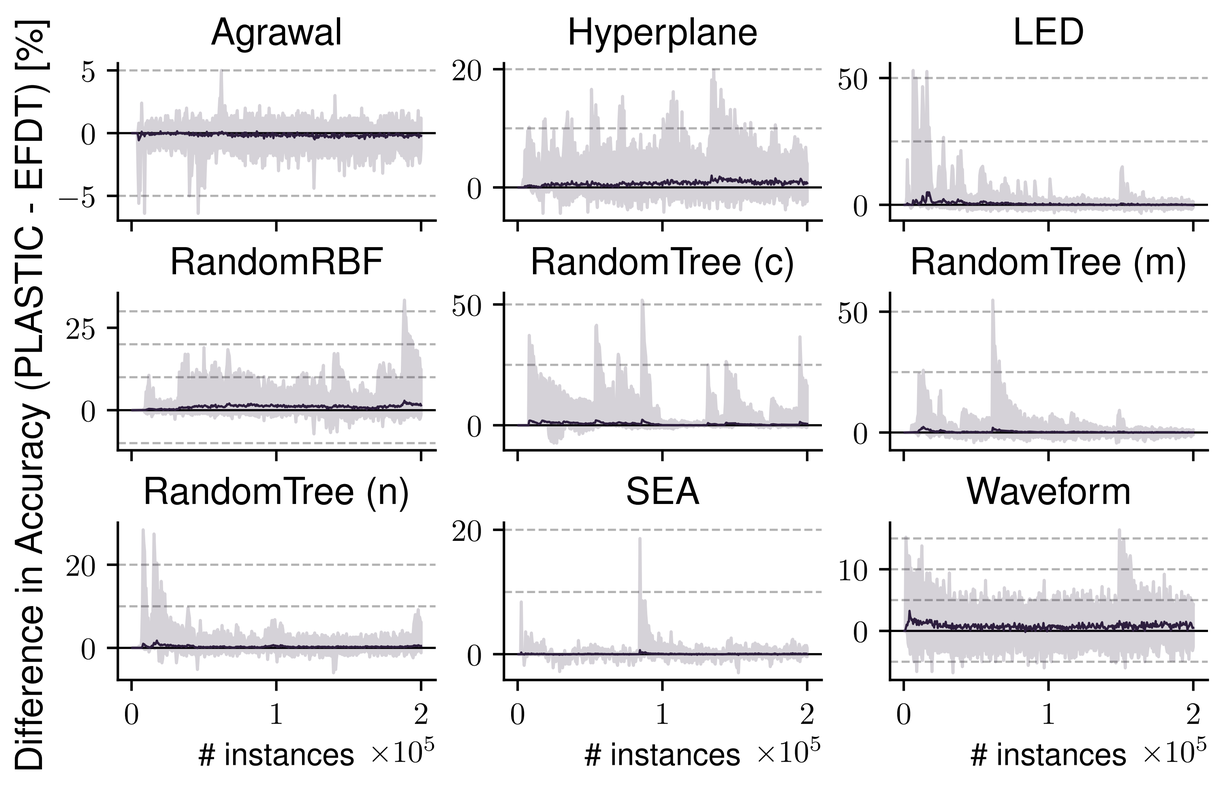
Algorithm

- Decouple decision tree branches
- Reorder each branch so that the desired split attribute is at root
- Rebuild subtree by re-connecting the branches



Results

- Graphs show the difference in accuracy between PLASTIC and EFDT
- Spike to the top → PLASTIC outperformed EFDT
- PLASTIC avoids EFDT's sudden accuracy drops caused by subtree pruning



There's more in the paper

- How we deal with numeric and binary splits
- Adaptive version of plastic, called PLASTIC-A
- Experimental results on real-world data streams
- Comparison against EFHAT [3], the state of the art approach for time-changing data streams

Visit our paper and github repository

Paper

https://doi.org/10.1007/978-3-031-70362-1_3

GitHub

https://github.com/heymarco/PLASTIC

References

[1] P. M. Domingos and G. Hulten, "Mining high-speed data streams," in *SIGKDD*, ACM, 2000, pp. 71–80. DOI: 10.1145/347090.347107.

[2] C. Manapragada, G. I. Webb, and M. Salehi, "Extremely fast decision tree," in *KDD*, ACM, 2018, pp. 1953–1962. DOI: 10.1145/3219819.3220005.

[3] C. Manapragada, M. Salehi, and G. I. Webb, "Extremely fast hoeffding adaptive tree," in *ICDM*, IEEE, 2022, pp. 319–328. DOI: 10.1109/ICDM54844.2022.00042.