THE OHIO STATE UNIVERSITY

Memory Mapping and Parallelizing Random Forests for Speed and Cache Efficiency

Eduardo Romero
romerogainza.1@osu.edu
The Ohio State University

Angela Li
li.10011@osu.edu
The Ohio State University

Christopher Stewart
cstewart@cse.ohio-state.edu
The Ohio State University

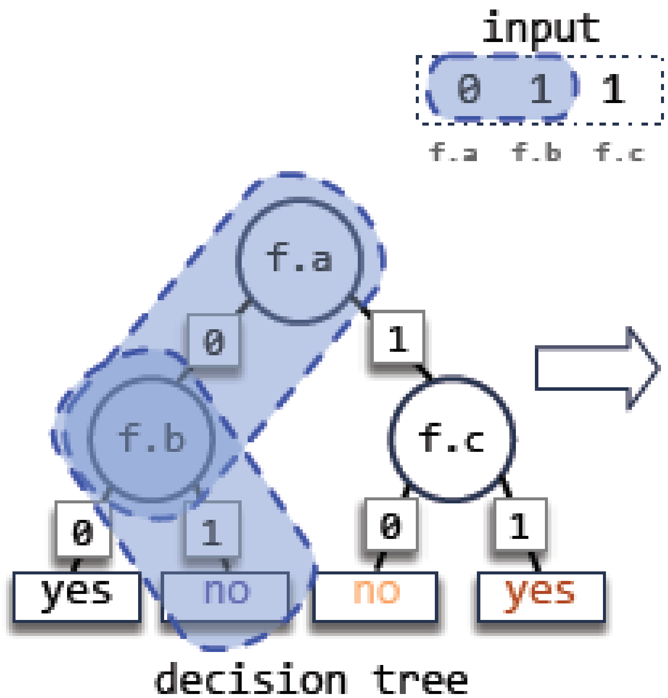
Kyle Hale
khale1@iit.edu
Illinois Institute of Technology

Nathaniel Morris
morris.743@osu.edu
The Ohio State University

Background – Memory Mapping

Memory mapping as a Depth-first alternative to tree traversal.

Table size grows exponentially



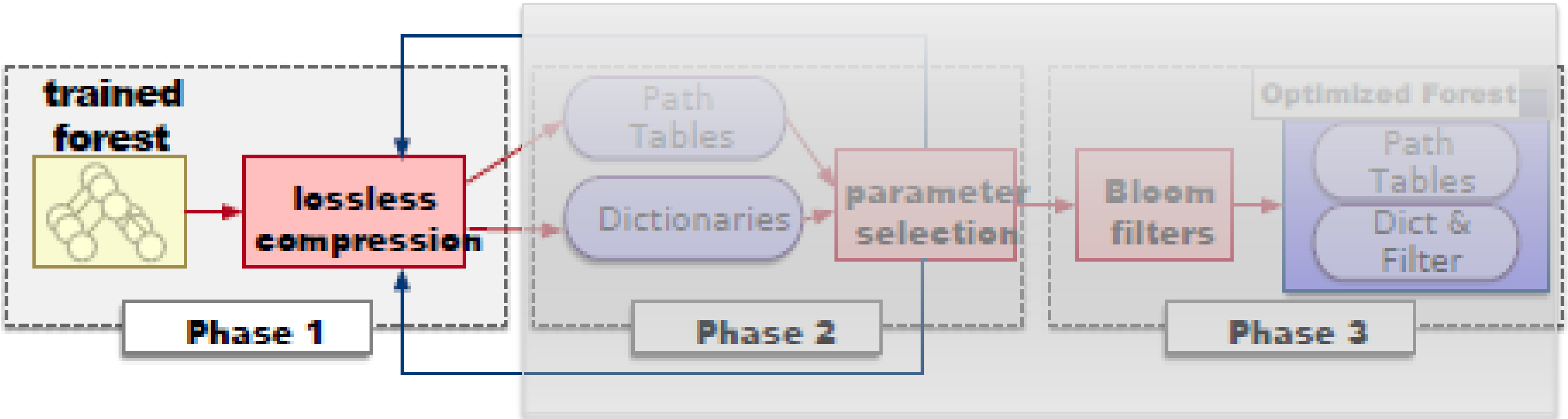
features form table address | table value

f.a	f.b	f.c	result
0	0	0	yes
0	0	1	yes
0	1	0	no
0	1	1	no
1	0	0	no
1	0	1	yes
1	1	0	no
1	1	1	yes

memory mapped table

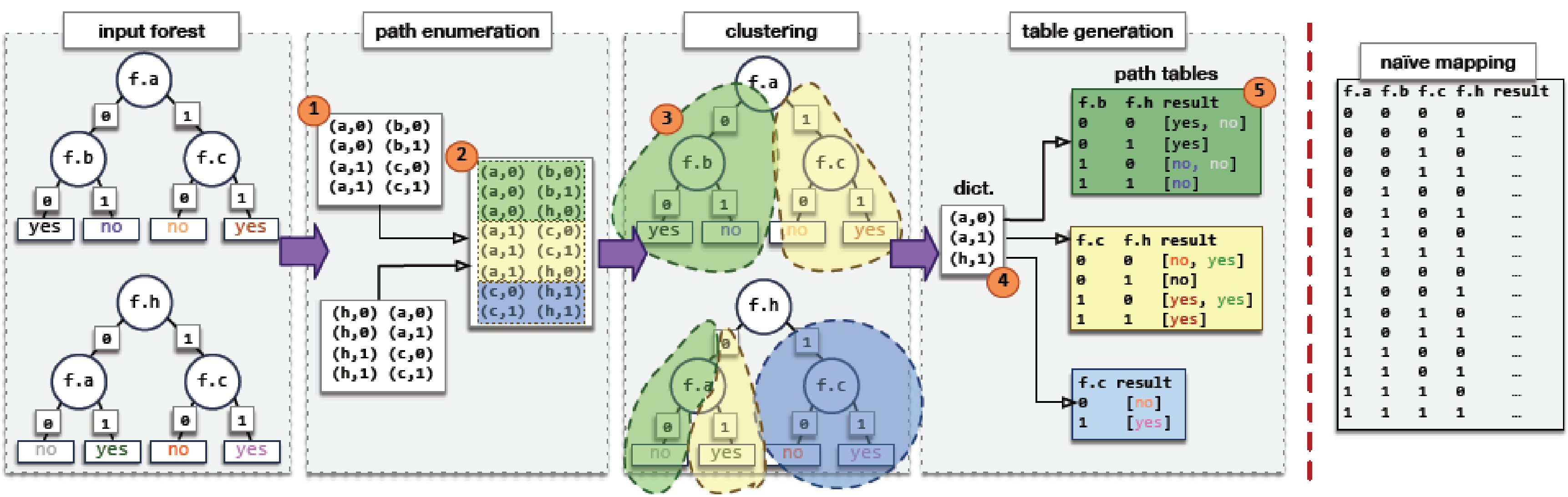
Design

Overview:



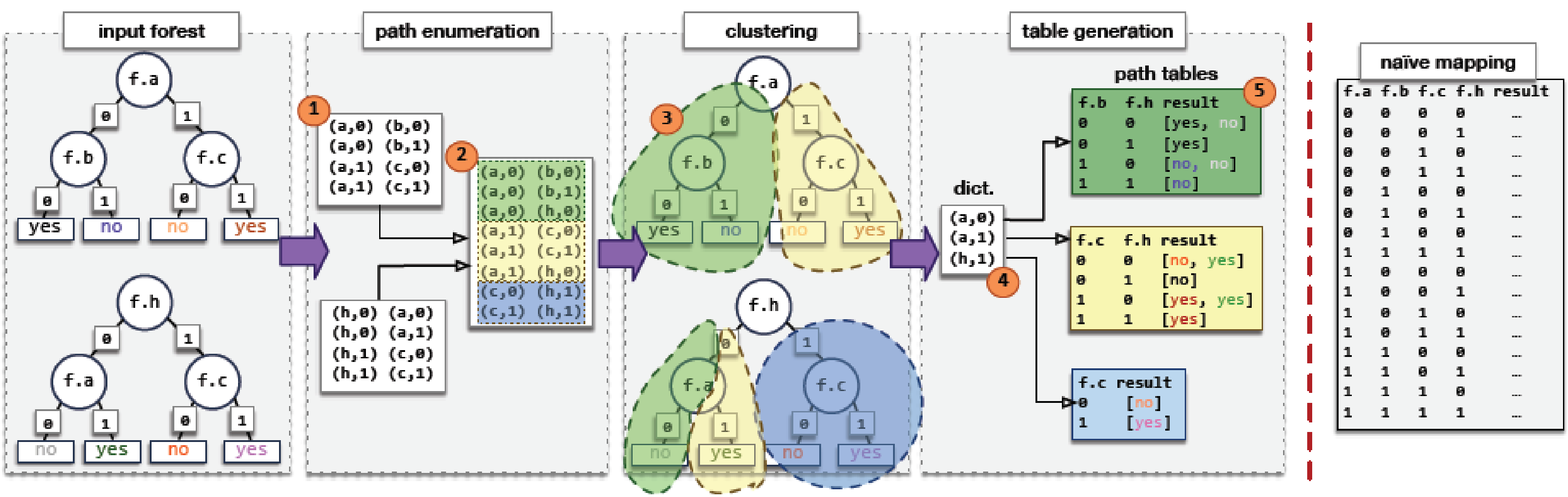
Phase 1 – Clustering & Compression

- List all paths in the forest.
- Sort features in each path.
- Sort all paths.



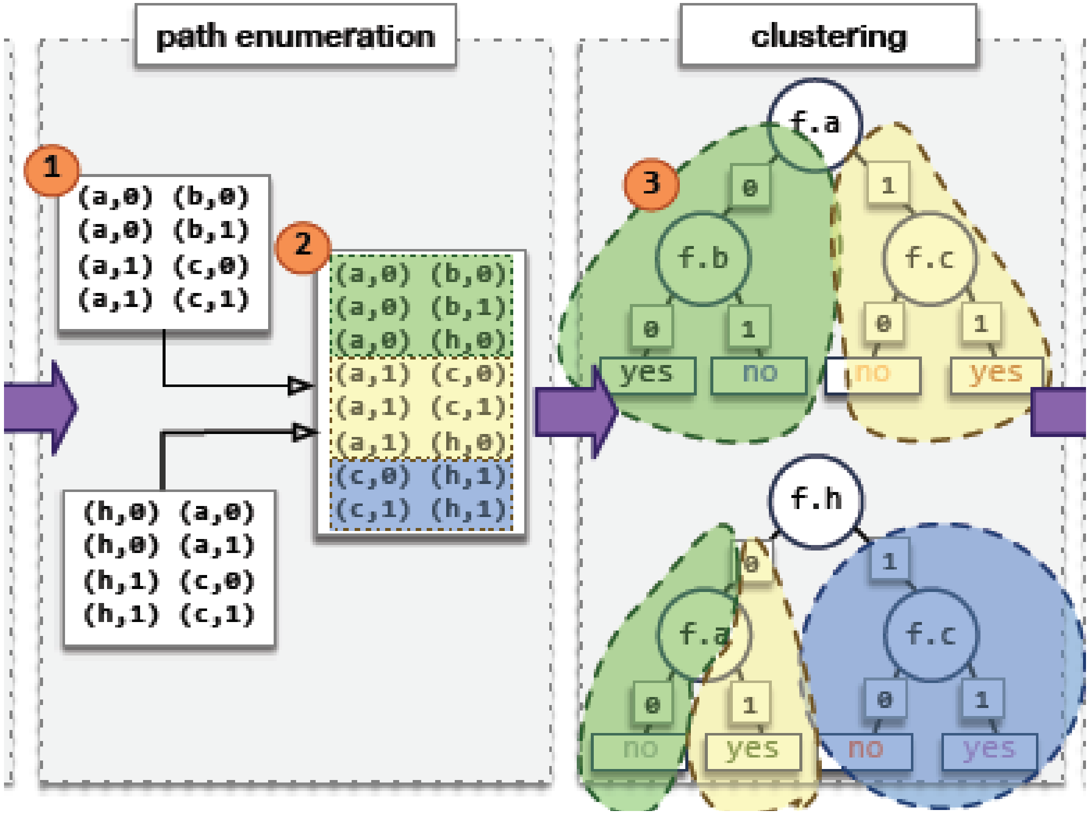
Phase 1 – Clustering & Compression

- Assign paths to clusters.
- Create table from each cluster
- Resulting table size < naïve table



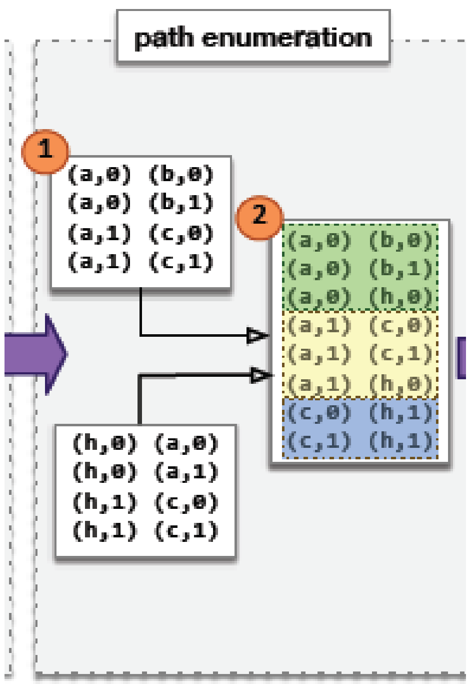
Phase 2: Parameter selection

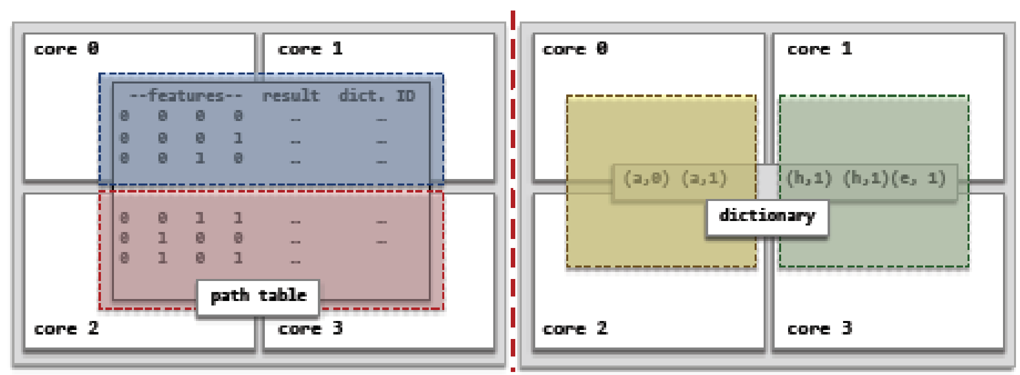
How are clusters divided?



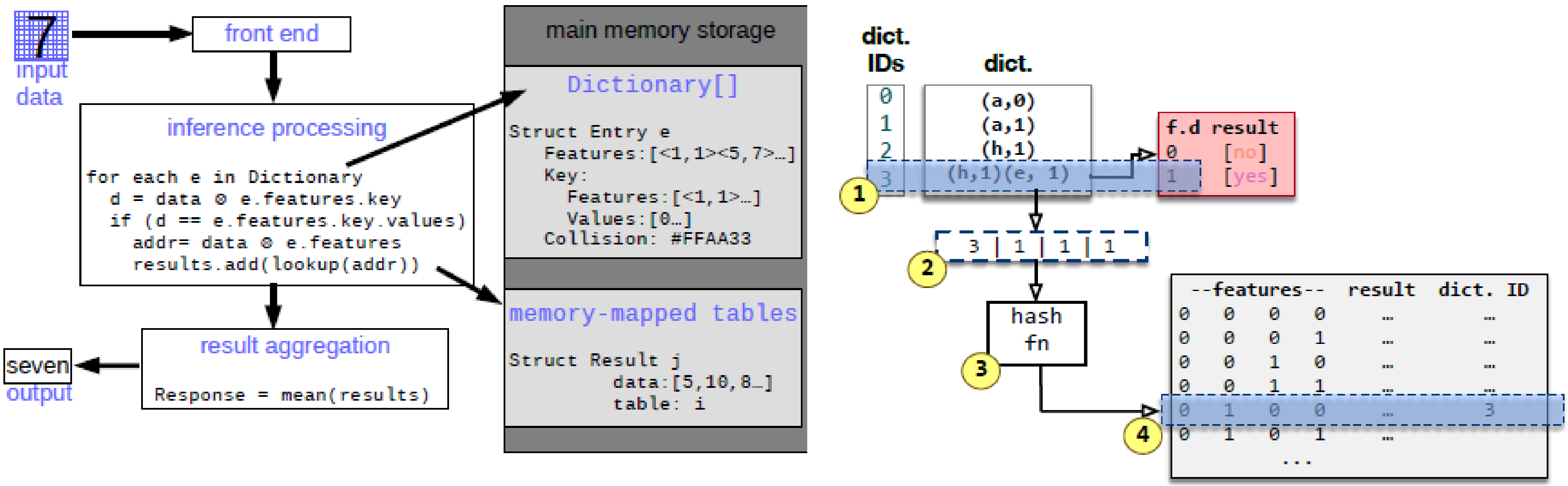
Phase 2: Parameter selection

- Tunable cluster size
- Assign tables and/or dictionaries to different clusters





Inference Overview

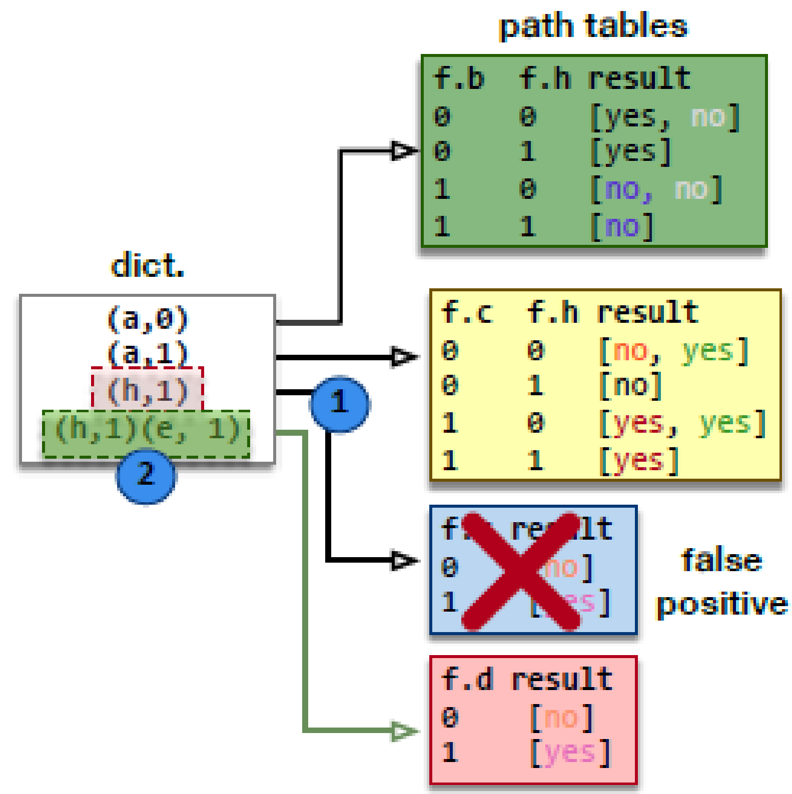


Phase 3: Improving Path Table selection

Could this model lead to false positives?
Yes. Hence Dictionary ID.

Bloom filter: Fast and Safe.

--features--				result	dict. ID
0	0	0	0	…	…
0	0	0	1	…	…
0	0	1	0	…	…
0	0	1	1	…	…
0	1	0	0	…	3
0	1	0	1	…	
...					



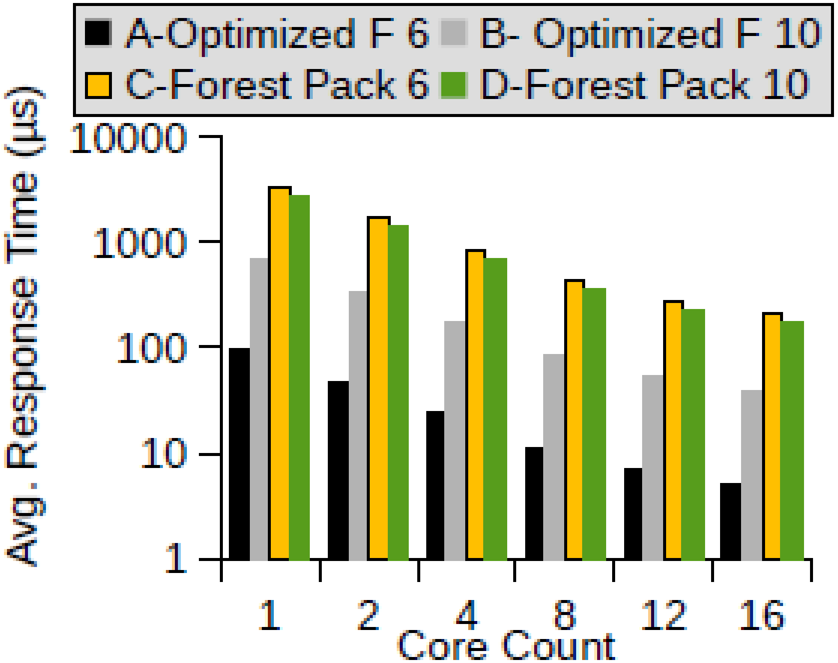
Early Results

MNIST Results:

Orders of magnitude faster than alternative approaches.

Multiple number of cores.

Two settings for max tree height.



256 trees parallelization response time (I)

Future Work

Search space methods

Optimization modeling

Explore datasets

Test interpretability benchmarks

References

J. Browne, D. Mhembere, T. M. Tomita, J. T. Vogelstein, and R. Burns. Forest packing: Fast parallel, decision forests. In Proceedings of the 2019 SIAM International Conference on Data Mining, pages 46–54. SIAM, 2019.