

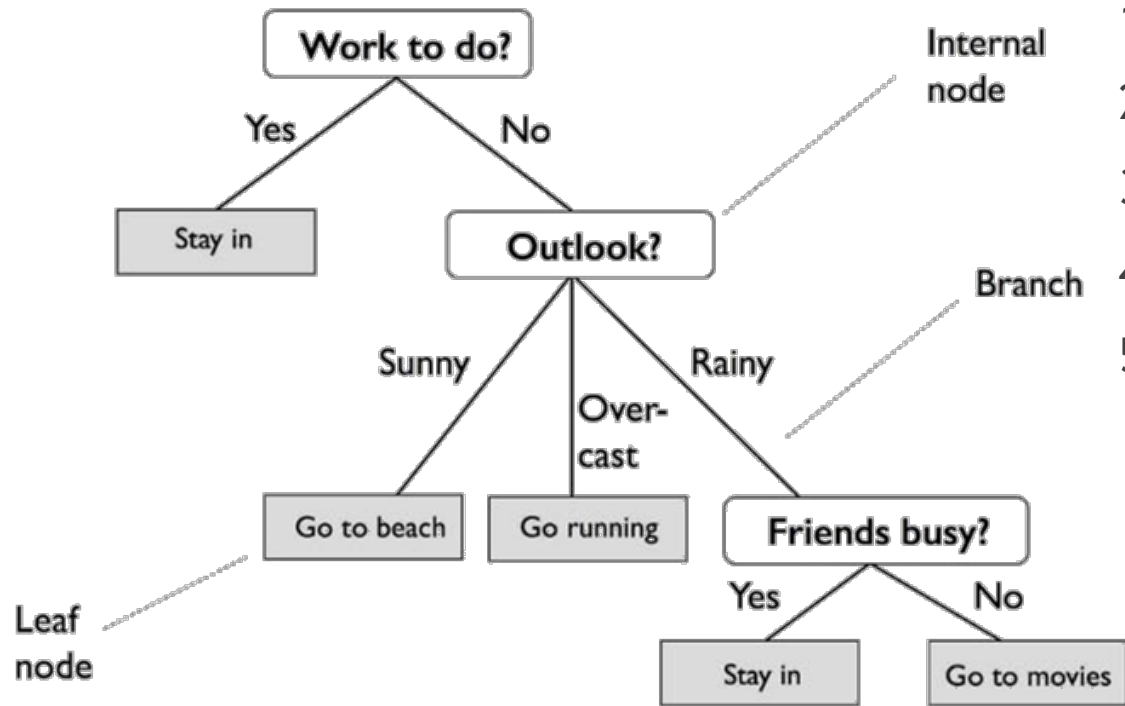
Approach to learning a decision tree with node relearning

Sergei Mitrofanov, Eugene Semenkin

Reshetnev Siberian State University of Science and Technology

Decision trees

Structure of decision tree

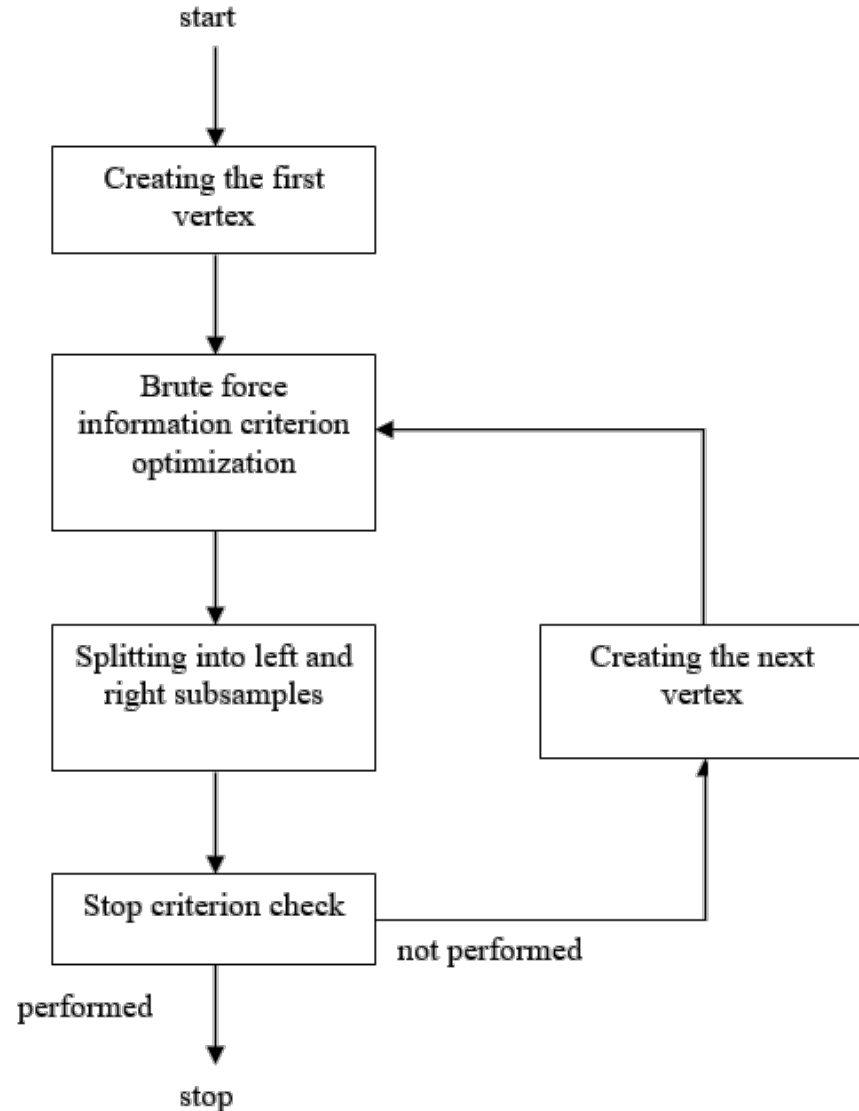


Settings and parameters

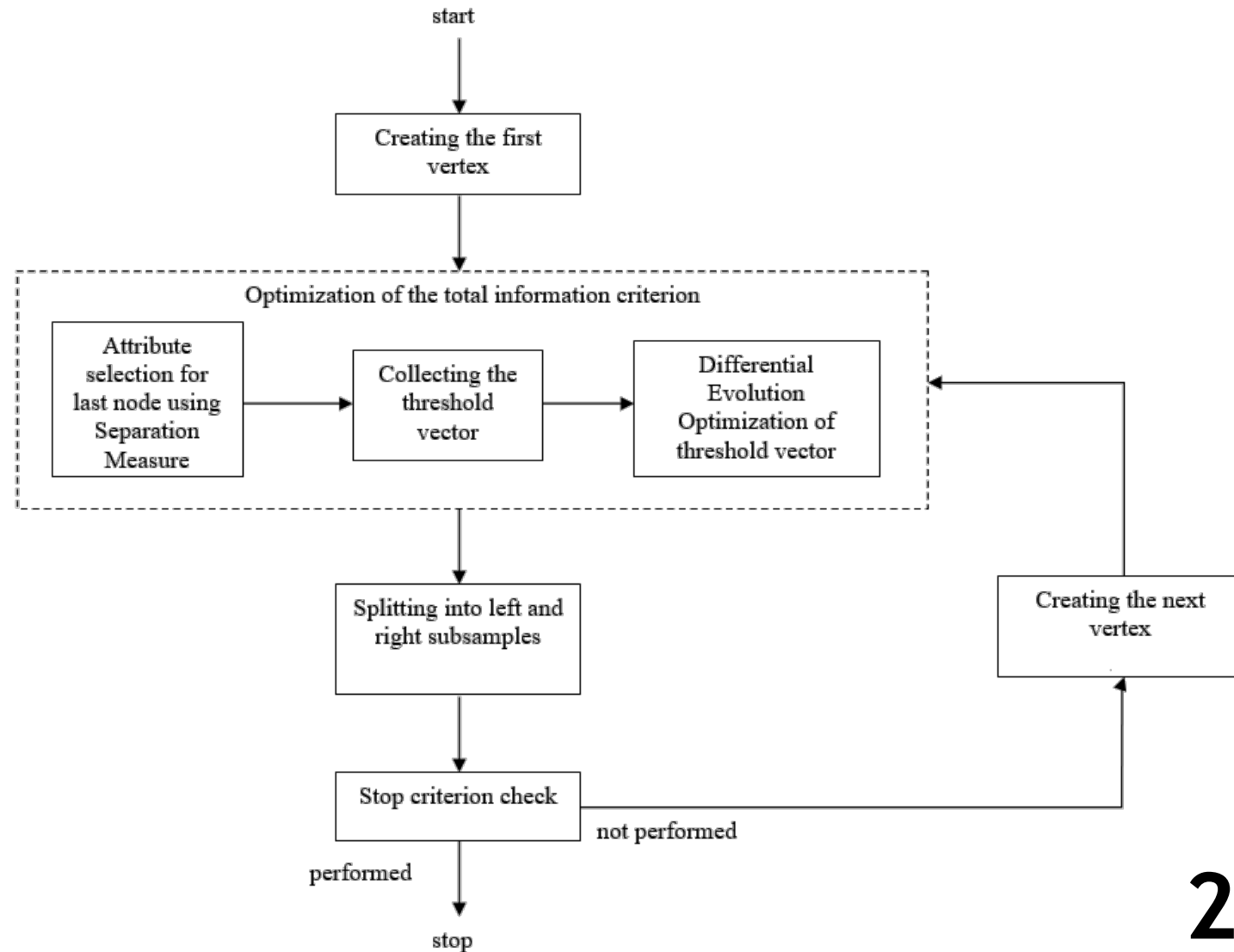
1. Types of predicates at the vertices;
2. Quality functional $Q(X, j, s)$;
3. Stop criterion;
4. Missing values processing method;
5. “Pruning” method.

Basic and new decision tree learning approaches

Basic decision tree learning algorithm



Modified decision tree learning algorithm

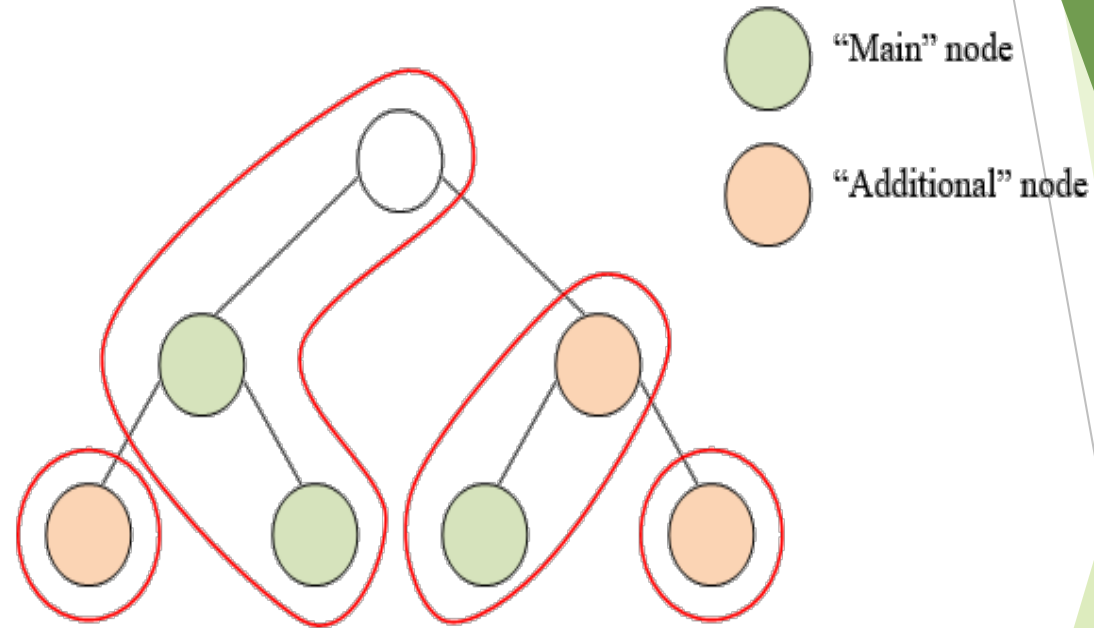


New learning approach

The threshold optimization will be done as follows:

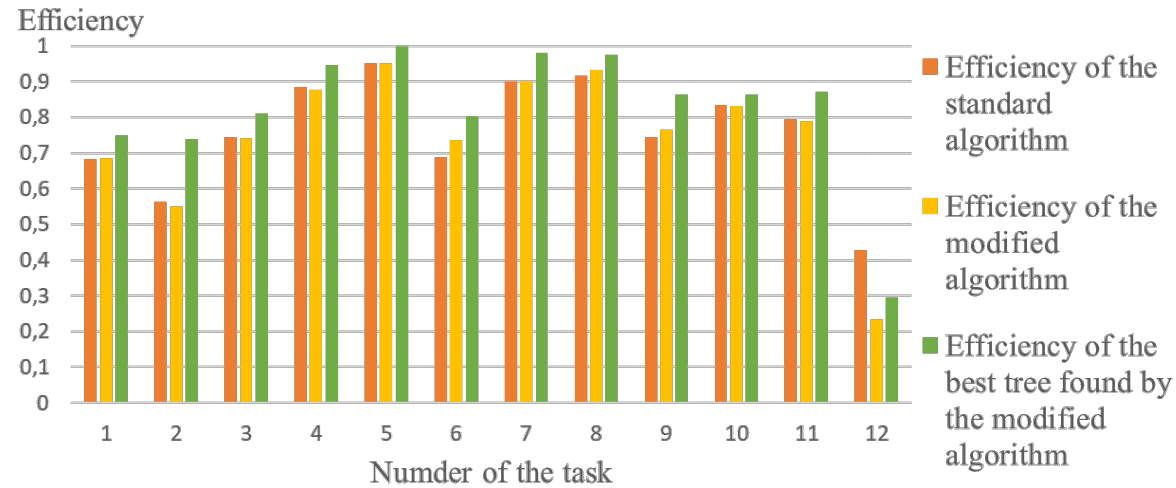
- 1) A threshold value is added to the vector from the learning node.
- 2) Let's go to step 3 in the case of considering the "main" node, otherwise we have finished the formation of the optimization vector.
- 3) Let's go to step 1, taking the parent node as a basis.

Threshold values added to the optimization vector are circled by a red line.

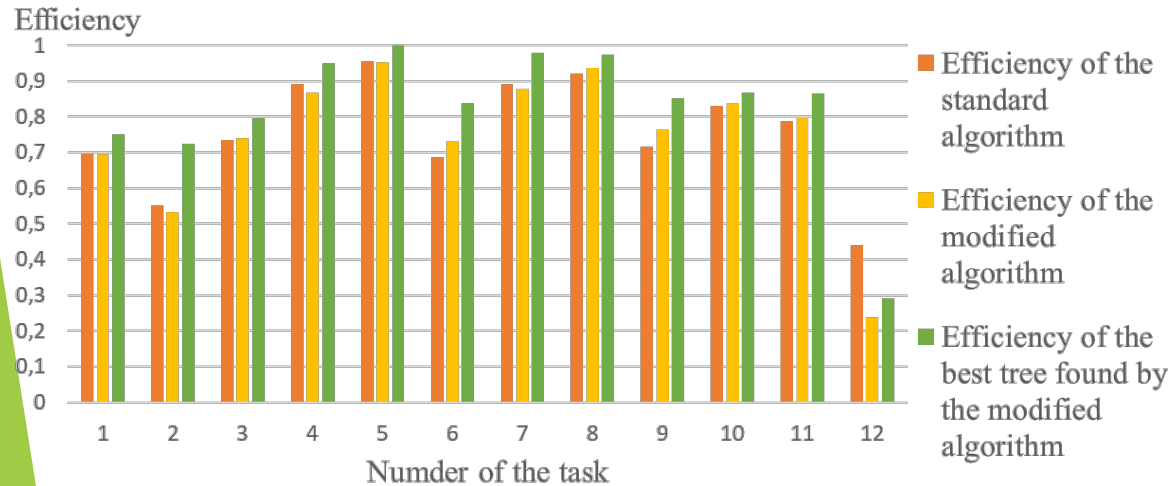


Selection of nodes for optimization

Solution of classification problems



Comparison of ID3 algorithms classification efficiency



Comparison of CART algorithms classification efficiency

Classification problems:

- 1) Determining the type of car.
- 2) Speaker accent recognition.
- 3) Determining the type of cityscape.
- 4) Recognizing the stage of hepatitis C.
- 5) Iris Variety Recognition.
- 6) Recognizing Parkinson's Disease.
- 7) Determination of the need for preventive maintenance of equipment.
- 8) Defining images by segments.
- 9) Recognizing a heart defect.
- 10) Soil type recognition from satellite imagery.
- 11) Recognition of biodegradable chemicals.
- 12) Workgroup classification by productivity in a garment factory.

Statistical analysis

ESTIMATED STUDENT T-TEST VALUES

Task number	ID3	CART
1	0.287	0.123
2	0.424	0.449
3	0.273	0.297
4	0.432	1.194
5	0.018	0.332
6	1.258	1.123
7	0.006	0.16
8	1.014	1.085
9	1.255	1.315
10	0.134	0.662
11	0.312	0.83
12	13.416	10.604

ESTIMATED STUDENT T-TEST VALUES (BEST VALUES)

Task number	ID3	CART
1	3.454	1.985
2	2.13	1.025
3	2.351	3.171
4	1.454	0.771
5	0.566	0
6	2.299	2.351
7	0.216	0.166
8	2.411	2.341
9	2.908	2.661
10	1.203	1.64
11	1.927	3.054
12	11.228	9.208

Conclusion

- ▶ The approach proposed in this article, on average, works the same as the standard algorithm.
- ▶ The modified algorithms sometimes allow finding decision trees that perform better on the task.
- ▶ In the future, it is planned to make a "smart" choice of an attribute for optimization and, possibly, change it during the training of the decision tree.