Tabular NAS

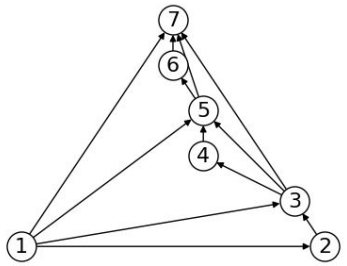
Vasily Ivanov, 2nd year PhD student

Selected topics in Data Science

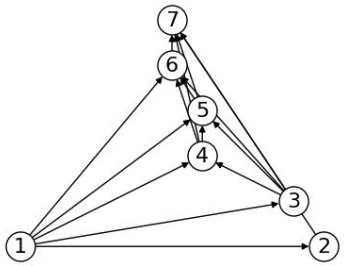
Skoltech

Differentiable NAS

Differentiable NAS is the way to obtain graph of the architecture from hypergraph using gradient descent



ASR cell



DARTS cell

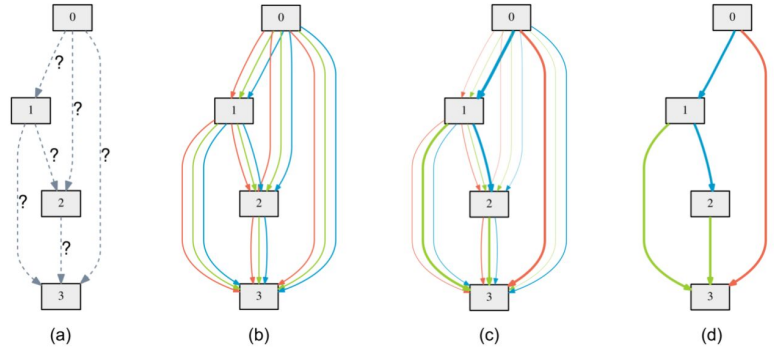


Figure 1: An overview of DARTS: (a) Operations on the edges are initially unknown. (b) Continuous relaxation of the search space by placing a mixture of candidate operations on each edge. (c) Joint optimization of the mixing probabilities and the network weights by solving a bilevel optimization problem. (d) Inducing the final architecture from the learned mixing probabilities.

(Liu et al., 2018)

Differentiable Nas Algorithms

Subsampling-like

GDAS
Proxyless NAS
etc

Train all at once (softmax)

DARTS-like

Goals

1. Try Differentiable Nas on tabular tasks
2. Test Modern Neural tabular solutions in NAS context

Methodology (Tabular Nas)

1. Select Tabular Datasets
2. Start from the tree baseline (Random Forest)
3. Design the Search Space
4. Experiment with search space
5. Test different NAS algorithms on this search space
6. Test attention-based and ResNet-like solutions

Datasets

5 biggest datasets for classification task in the NIPS tabular paper (link in report)

Dataset	Folder Name	Type	Num Features	Train Size
Covertype	covtype	multiclass	54	371847
Higgs Small	higgs-small	binclass	28	62751
Otto Group Products	otto	multiclass	93	39601
Adult	adult	binclass	7	26048
Churn Modelling	churn	binclass	10	6400

Table 1: Properties of used datasets

Results

method	Covertype	Higgs	Otto	Adult	Churn Modelling
random forest	82.6992	71.3870	75.6948	84.2945	85.3500
Transformer_op	36.46	52.856	13.009	76.377	79.65
Darts_cell_op	-	52.856	72.697	76.377	79.65
ASR_cell_op	-	47.144	65.999	76.377	79.65

Table 2: Accuracy for the experiments with baseline rf, different cells, darts, transformers

algorithm	Covertype	Higgs	Otto	Adult	Churn Modelling
Darts	-	47.144	65.999	76.377	79.65
GDAS	-	47.144	26.05	76.384	79.65
DrNAS	-	52.856	57.482	76.377	79.65

Table 3: Accuracy for the different algorithms

Conclusions

1. Differentiable NAS algorithms are very unstable
2. NASLIB barely works or doesn't work at all
3. I had some good results but couldn't reproduce them
4. Seemed like simple DARTS worked better of them all
5. My current committed / reproducible results suggest that DrNAS results were better but I don't quite believe it

Thank you for attention