Improving Deep Forest by Exploiting High-order Interactions

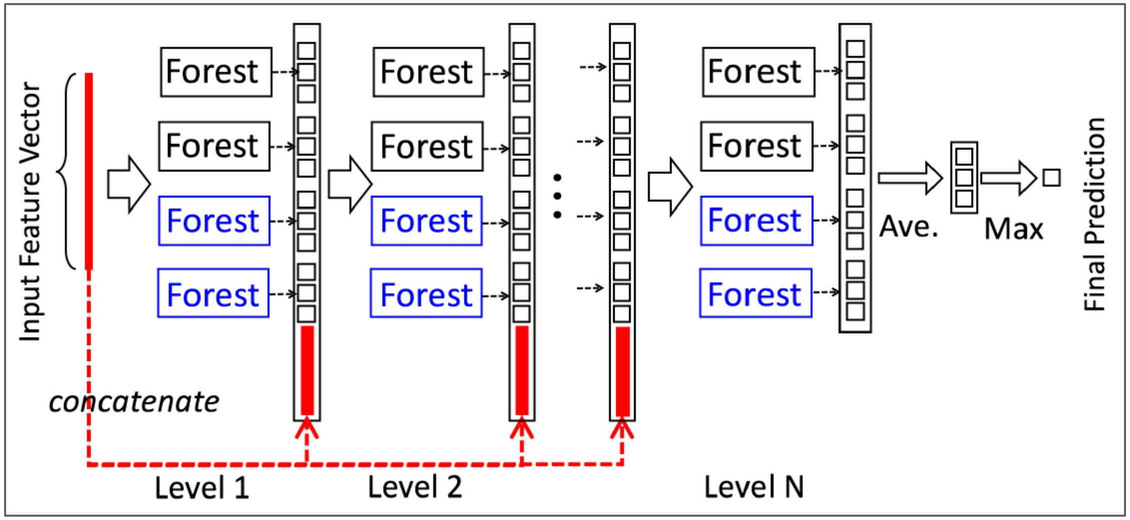
Yi-He Chen* Shen-Huan Lyu* Yuan Jiang†

{ chenyh, lvsh, jiangy }@lamda.nju.edu.cn

* : equal contribution
† : corresponding author

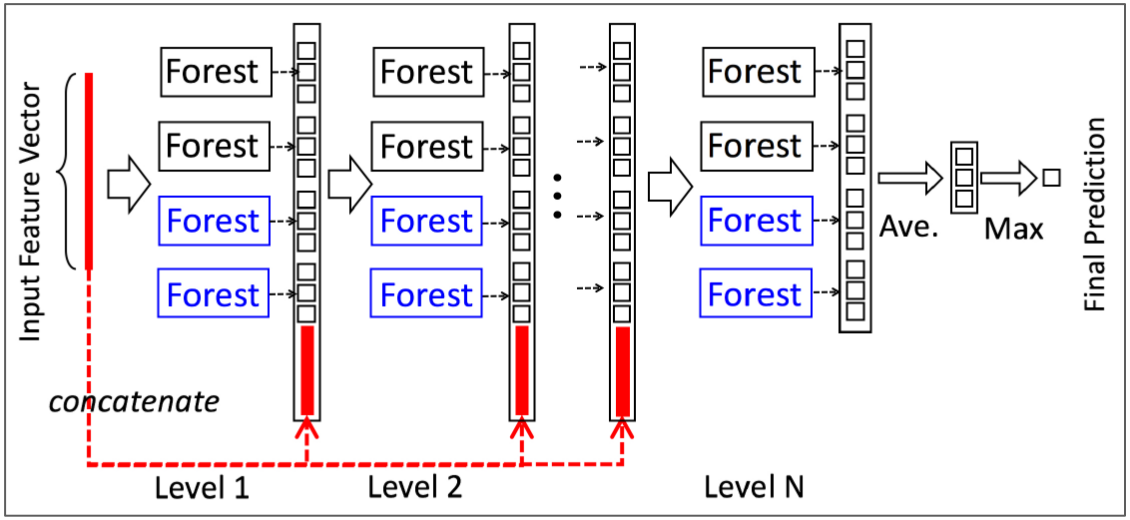
Deep Forest

- Realize deep learning with non-differentiable modules
- Excellent performance on categorical & mixed modeling tasks



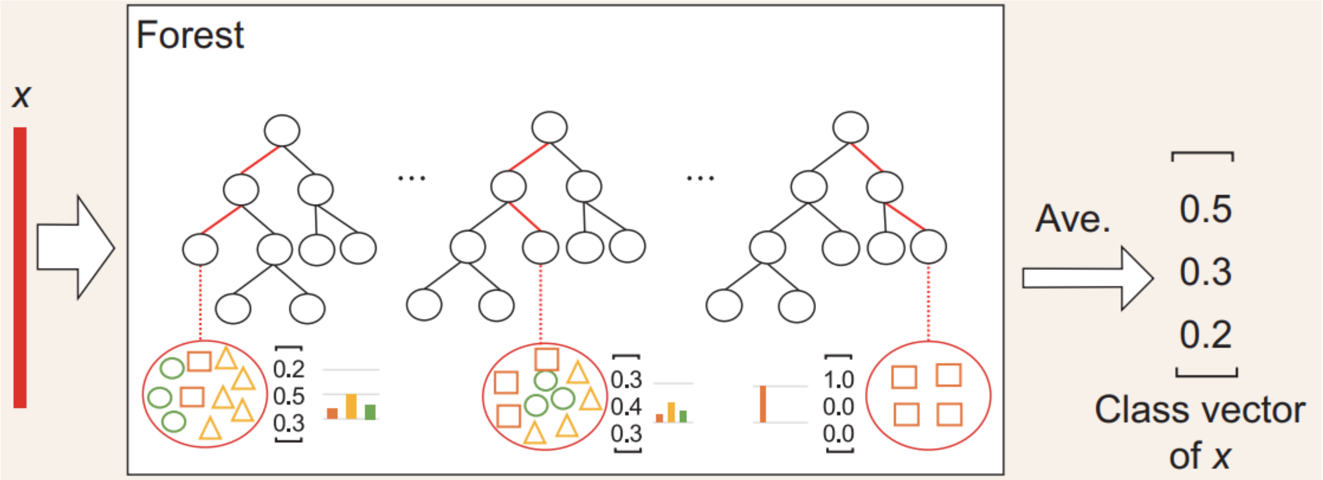
Future **Issues** of Deep Forest :

- Representation learning ability and diversity are limited.
- High memory and time consumption.



Future **Issues** of Deep Forest :

- Representation learning ability and diversity are limited.
- High memory and time consumption.



Main idea:

Utilize **decision paths** of decision trees inside random forest.

Main idea:

Utilize **decision paths** of decision trees inside random forest.

decision paths = decision rules = leaf nodes

$$\boxed{\text{IF: } condition_1 \ \& \ldots \& \ condition_i, \ \text{THEN: } response.}$$

Main idea:

Utilize **decision paths** of decision trees inside random forest.

decision paths = decision rules = leaf nodes

$$\boxed{\text{IF: } condition_1 \;\& \ldots \& \; condition_i, \text{ THEN: } response.}$$

Extract decision paths from RF?

Main idea:

Utilize **decision paths** of decision trees inside random forest.

decision paths = decision rules = leaf nodes

$$\boxed{\text{IF: } condition_1 \ \& \ldots \& \ condition_i, \ \text{THEN: } response.}$$

Extract decision paths from RF?

No!

1. Lack of statistical importance.
2. Complex decision rules do not generalize well.

Random Intersection Trees

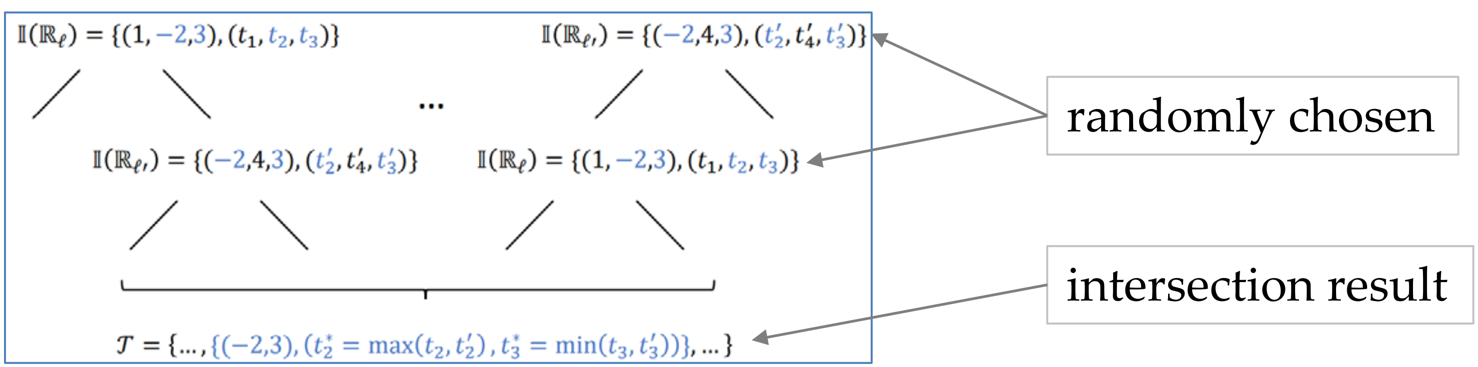
--- start with the full set of variables as an interaction and then iteratively prune away variables.

Random Intersection Trees

--- start with the full set of variables as an interaction and then iteratively prune away variables.

Perform **intersection** among random chosen feature interactions.

(intersection of decision rules == union of corresponding regions)



Utilizing discovered feature interactions

--- Activated Linear Combination (ALC)

$$\boldsymbol{r} = \{\ldots, \mathrm{ReLU}\big(-w_2(x_2 - t_2^*) + w_3(x_3 - t_3^*)\big), \ldots\}$$

Learning And Mining from DatA

Utilizing discovered feature interactions

--- Activated Linear Combination (ALC)

$$\boldsymbol{r} = \{\dots, \mathrm{ReLU}\big(-w_2(x_2 - t_2^*) + w_3(x_3 - t_3^*)\big), \dots\}$$

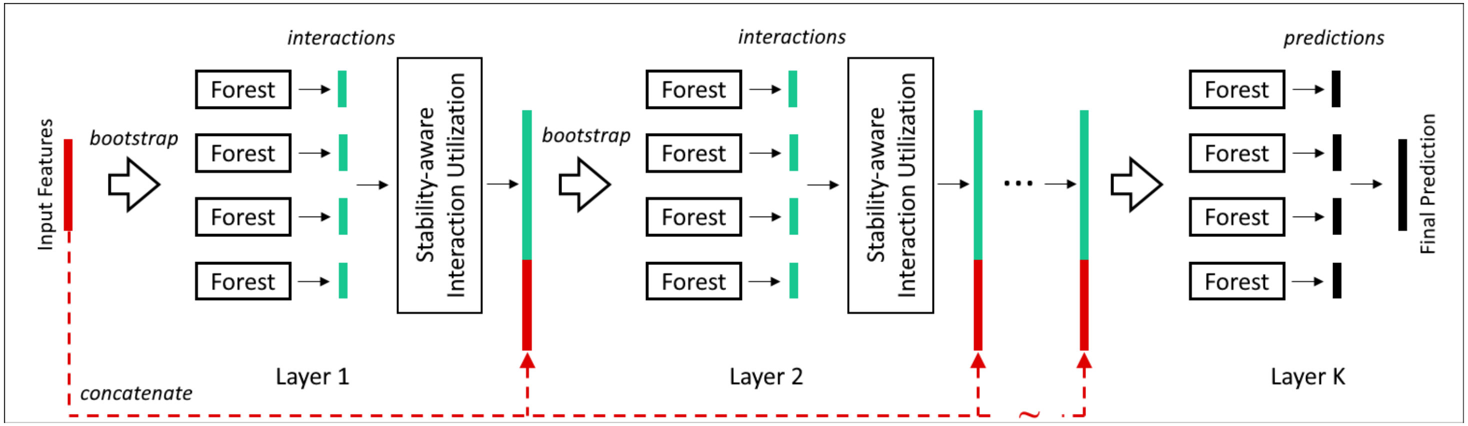
$\Downarrow$

Distributed Representation

--- core idea of deep learning methods.

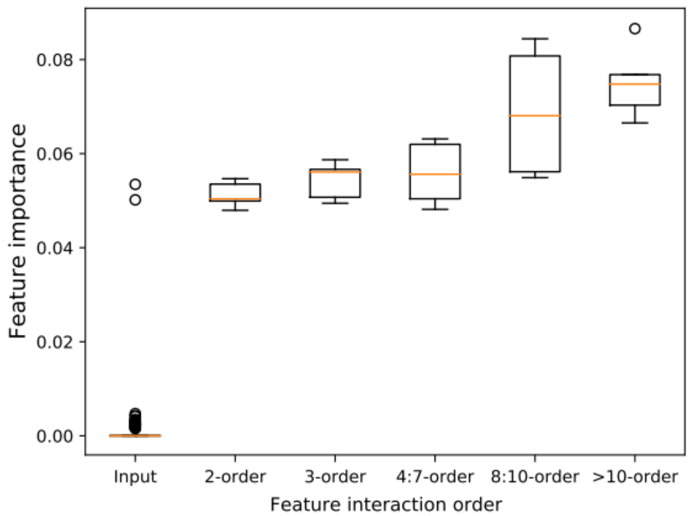
Adaptive layer growth

--- retains high-order interactions.



Adaptive layer growth

--- retains high-order interactions.



The importance of features with varying interaction orders

Experiments : UCI data

Datasets	hiDF	gcForest	$\text{gcForest}_{\text{CS}}$	Random Forests	XGBoost	GBDT	SVM
CoverType	**97.62 ± 0.08**	96.23 ± 0.10	96.09 ± 0.07	95.63 ± 0.07	94.48 ± 0.05	96.94 ± 0.11 •	71.47 ± 0.12
Adult	**86.90 ± 0.05**	86.17 ± 0.06	86.17 ± 0.09	85.15 ± 0.08	86.40 ± 0.00 •	86.12 ± 0.10	79.86 ± 0.00
Bank	**89.96 ± 0.21**	89.89 ± 0.26	89.95 ± 0.25 •	89.33 ± 0.17	89.13 ± 0.33	88.60 ± 0.25	89.75 ± 0.21
Credit Card	**82.00 ± 0.23**	81.74 ± 0.38 •	81.73 ± 0.20	81.71 ± 0.33	80.61 ± 0.36	81.24 ± 0.27	77.88 ± 0.01
Arrhythmia	**78.24 ± 1.62**	76.26 ± 1.49 •	74.51 ± 1.62	74.51 ± 2.13	75.16 ± 1.32	74.29 ± 3.15	60.66 ± 1.62
YearPredictionMSD	**75.89 ± 0.46**	75.49 ± 0.40	75.61 ± 0.30	73.41 ± 0.30	74.69 ± 0.21	75.85 ± 0.20 •	68.62 ± 0.62
Diabetes	**62.43 ± 0.23**	62.26 ± 0.36	62.37 ± 0.33 •	62.11 ± 0.24	61.00 ± 0.17	61.04 ± 0.45	54.99 ± 0.02
Satimage	**91.75 ± 0.06**	91.63 ± 0.12 •	91.58 ± 0.11	91.21 ± 0.05	90.65 ± 0.00	90.44 ± 0.09	88.60 ± 0.00
Crowdsourced Mapping	**65.07 ± 0.88**	64.93 ± 0.95	65.03 ± 0.56 •	63.47 ± 0.72	62.00 ± 0.00	62.53 ± 0.50	55.67 ± 0.00
Congestive Heart Failure	**90.16 ± 0.53**	88.61 ± 0.59	88.58 ± 0.27	87.90 ± 0.01	89.34 ± 0.01 •	88.83 ± 0.01	85.09 ± 0.00
win/tie/lose	—	10/0/0	10/0/0	10/0/0	10/0/0	10/0/0	10/0/0

Experiments : UCI data

Datasets	hiDF	gcForest	$gcForest_{CS}$	Random Forests	XGBoost	GBDT	SVM
CoverType	**97.62** $\pm$ **0.08**	96.23 $\pm$ 0.10	96.09 $\pm$ 0.07	95.63 $\pm$ 0.07	94.48 $\pm$ 0.05	96.94 $\pm$ 0.11 $\bullet$	71.47 $\pm$ 0.12
Adult	**86.90** $\pm$ **0.05**	86.17 $\pm$ 0.06	86.17 $\pm$ 0.09	85.15 $\pm$ 0.08	86.40 $\pm$ 0.00 $\bullet$	86.12 $\pm$ 0.10	79.86 $\pm$ 0.00
Bank	**89.96** $\pm$ **0.21**	89.89 $\pm$ 0.26	89.95 $\pm$ 0.25 $\bullet$	89.33 $\pm$ 0.17	89.13 $\pm$ 0.33	88.60 $\pm$ 0.25	89.75 $\pm$ 0.21
Credit Card	**82.00** $\pm$ **0.23**	81.74 $\pm$ 0.38 $\bullet$	81.73 $\pm$ 0.20	81.71 $\pm$ 0.33	80.61 $\pm$ 0.36	81.24 $\pm$ 0.27	77.88 $\pm$ 0.01
Arrhythmia	**78.24** $\pm$ **1.62**	76.26 $\pm$ 1.49 $\bullet$	74.51 $\pm$ 1.62	74.51 $\pm$ 2.13	75.16 $\pm$ 1.32	74.29 $\pm$ 3.15	60.66 $\pm$ 1.62
YearPredictionMSD	**75.89** $\pm$ **0.46**	75.49 $\pm$ 0.40	75.61 $\pm$ 0.30	73.41 $\pm$ 0.30	74.69 $\pm$ 0.21	75.85 $\pm$ 0.20 $\bullet$	68.62 $\pm$ 0.62
Diabetes	**62.43** $\pm$ **0.23**	62.26 $\pm$ 0.36	62.37 $\pm$ 0.33 $\bullet$	62.11 $\pm$ 0.24	61.00 $\pm$ 0.17	61.04 $\pm$ 0.45	54.99 $\pm$ 0.02
Satimage	**91.75** $\pm$ **0.06**	91.63 $\pm$ 0.12 $\bullet$	91.58 $\pm$ 0.11	91.21 $\pm$ 0.05	90.65 $\pm$ 0.00	90.44 $\pm$ 0.09	88.60 $\pm$ 0.00
Crowdsourced Mapping	**65.07** $\pm$ **0.88**	64.93 $\pm$ 0.95	65.03 $\pm$ 0.56 $\bullet$	63.47 $\pm$ 0.72	62.00 $\pm$ 0.00	62.53 $\pm$ 0.50	55.67 $\pm$ 0.00
Congestive Heart Failure	**90.16** $\pm$ **0.53**	88.61 $\pm$ 0.59	88.58 $\pm$ 0.27	87.90 $\pm$ 0.01	89.34 $\pm$ 0.01 $\bullet$	88.83 $\pm$ 0.01	85.09 $\pm$ 0.00
win/tie/lose	—	10/0/0	10/0/0	10/0/0	10/0/0	10/0/0	10/0/0

Computational complexity

Datasets	Method	Training time	Test time	Memory
Adult	hiDF	3208.7	13.1	2793.2
	gcForest	2763.8	541.3	29243.9
	$gcForest_{CS}$	1063.1	290.2	23272.7
Covertype	hiDF	4505.9	90.9	5562.5
	gcForest	2861.3	1519.4	61981.2
	$gcForest_{CS}$	1582.3	730.1	57658.8

Summary :

Interaction-based representation. & Multi-layer structure.

--- realize hierarchical distributed representation.

--- enhance representation diversity.

--- reduce computational complexity.

Thanks!