C^*-algebra Net: A New Approach Generalizing Neural Network Parameters to C^*-algebra

Yuka Hashimoto[1], Zhao Wang[1,2], and Tomoko Matsui[3]

1. NTT Network Service Systems Laboratories, NTT Corporation
2. Institute for Disaster Response Robotics, Waseda University
3. Department of Statistical Modeling, the Institute of Statistical Mathematics

Motivation: Learning multiple neural networks

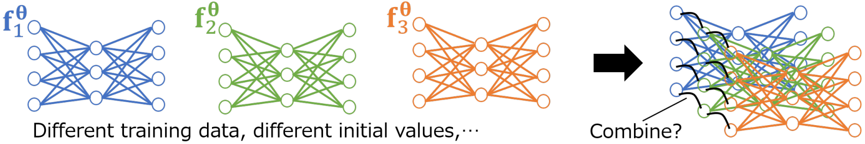
Standard neural network:

$$\mathbf{f}^{\boldsymbol{\theta}} = \boldsymbol{\sigma}_{H+1} \circ \mathbf{W}_{H+1} \circ \boldsymbol{\sigma}_H \circ \mathbf{W}_H \circ \cdots \circ \boldsymbol{\sigma}_1 \circ \mathbf{W}_1$$

$N_0, \ldots, N_{H+1}$: dimension of each layer,
$\mathbf{W}_i : \mathbb{R}^{N_{i-1}} \to \mathbb{R}^{N_i}$ $(i = 1, \ldots, H+1)$: $N_{i-1} \times N_i$ matrix,
$\boldsymbol{\sigma}_i : \mathbb{R}^{N_i} \to \mathbb{R}^{N_i}$: (often nonlinear) activation function, $\boldsymbol{\theta} = [\mathbf{W}_i]$

When we need to learn multiple related networks, how can we learn them efficiently?

One solution: Generalize the neural network parameters to C^*-algebra

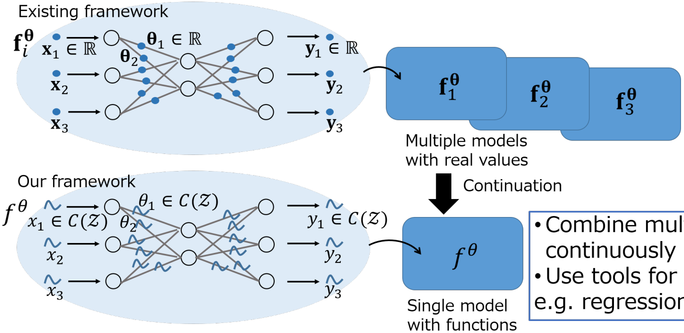


C^*-algebra network

C^*-algebra network:

$$f^\theta = \sigma_{H+1} \circ W_{H+1} \circ \sigma_H \circ W_H \circ \cdots \circ \sigma_1 \circ W_1$$

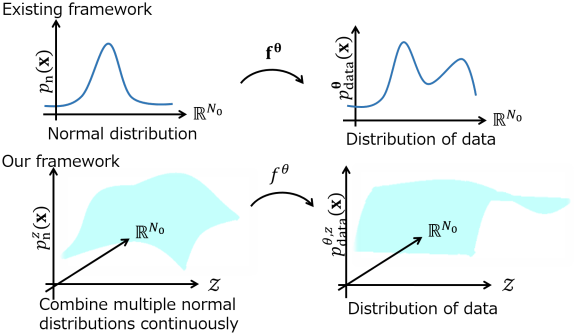
$\mathcal{A} = C(\mathcal{Z})$: C^*-algebra of continuous functions,
$W_i : \mathcal{A}^{N_{i-1}} \to \mathcal{A}^{N_i}$ $(i = 1, \ldots, H+1)$: $N_{i-1} \times N_i$ $\mathcal{A}$-valued matrix,
$\sigma_i : \mathcal{A}^{N_i} \to \mathcal{A}^{N_i}$: (often nonlinear) activation function, $\theta = [W_i]$



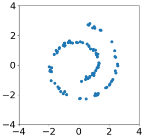
- Combine multiple models continuously
- Use tools for functions e.g. regression, integration

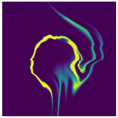
Application to density estimation

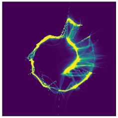
Normalizing flow:

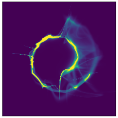


Numerical results









Training samples	Existing, single model	Existing, multiple models	Ours
NLL (avg.±std. dev.):	3.91 ± 0.621	4.03 ± 0.268	**2.94±0.057**

Other applications

Examples of other applications:

- Few-shot learning
- Ensemble learning
- Generating time-series or spatial data
- Learning distributions of parameters
- Generalizing complex-valued networks

Our framework is valid for a wide range of applications!

Email: yuka.hashimoto.rw@hco.ntt.co.jp