Noise-Robustness Through Noise: A Framework combining Asymmetric LoRA with Poisoning MoE

NeruIPS 2025

NEURAL INFORMATION PROCESSING SYSTEMS

NeruIPS 2025 Poster

1.School of Information and Software Engineering,
University of Electronic Science and Technology of China

2.School of Computer Science and Engineering, Central South University

Noise-Robustness Through Noise: A Framework combining Asymmetric LoRA with Poisoning MoE

Zhaokun Wang[1], Jinyu Guo[1*], Jingwen Pu[1], Lingfeng Chen[1],
Hongli Pu[1], Jie Ou[1], Libo Qin[2], Wenhong Tian[1*]
[1]School of Information and Software Engineering,
University of Electronic Science and Technology of China
[2]School of Computer Science and Engineering, Central South University

CONTENT

Background

Pre-trained language models (PrLMs) have demonstrated remarkable success across various NLP tasks.

To enhance model performance on downstream tasks, researchers typically employ domain-specific corpora for targeted **fine-tuning** of pre-trained models.

Background

In downstream NLP applications, noise poses multiple critical challenges:

1) labeling errors, syntactic irregularities, and extraneous content;

2) noise weakens model generalization when encountering unseen data;

3) noisy data may introduce biases.

Introduction

Recent reasearch about data noise:

1) focuses on reconstructing the data before training by cleaning, filtering, or relabeling to construct purified datasets;

2) involves developing dedicated denoising architectures during training.

Introduction

Two primary limitations in current research paradigms:

1) heavily rely on **manual intervention** or **prior assumptions** during data pre-processing, requiring noise detection and cleaning before model training;

2) focus on **improving the model architecture** during training avoid explicit data cleaning, but still **cannot avoid discriminating noise information.**

Introduction

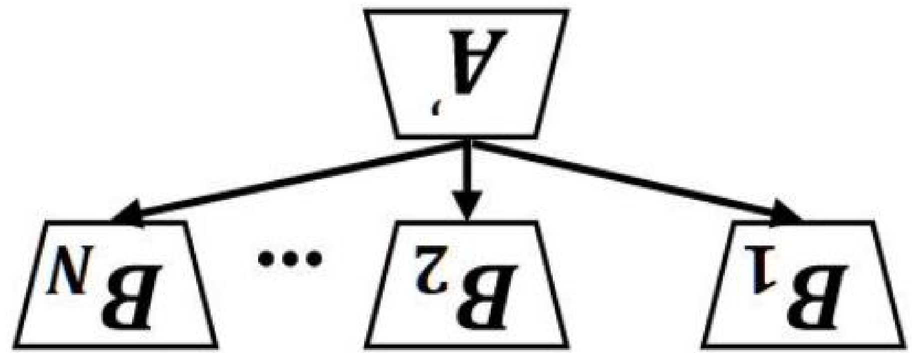
Thinking

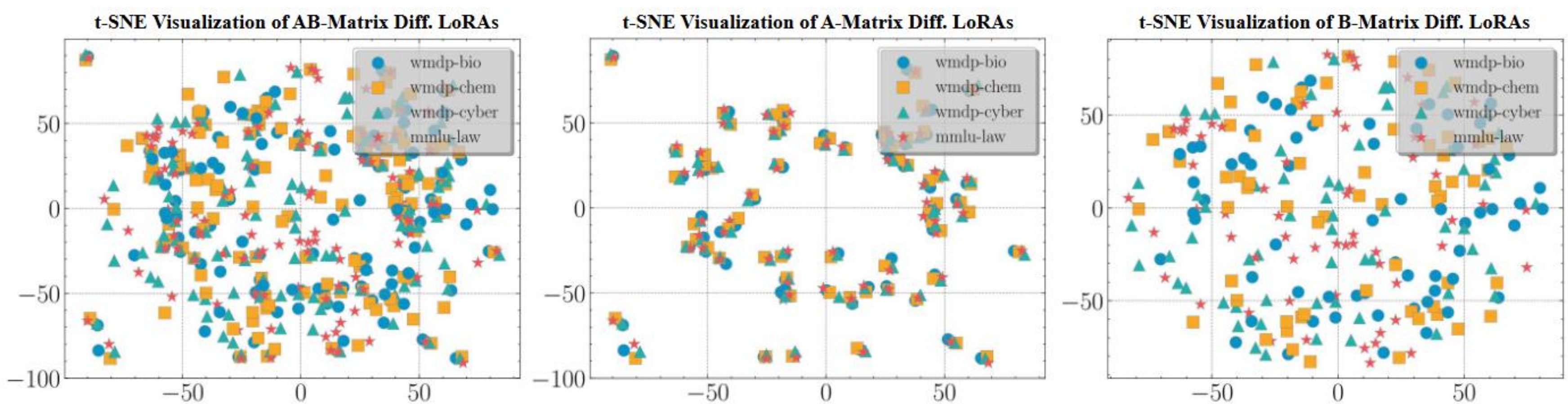
data processing： additional computational and annotation expenses but also lead to error propagation.

Noise injection： cost-effective and easily automatable

Introduction

Mixture-of-Experts



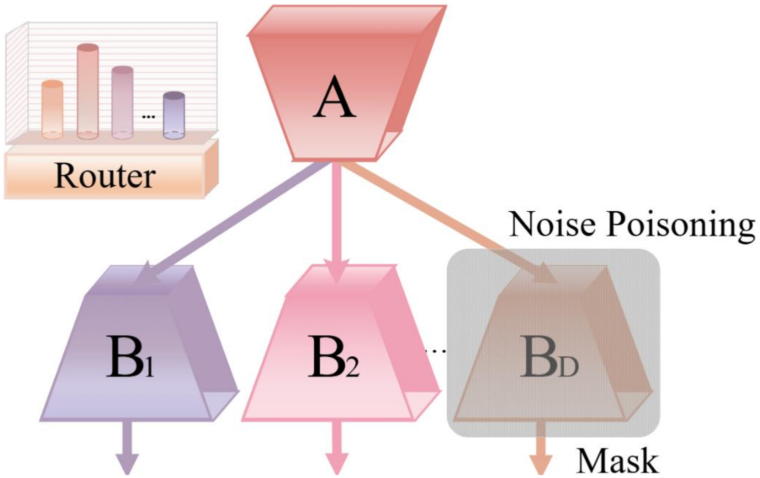


Introduction

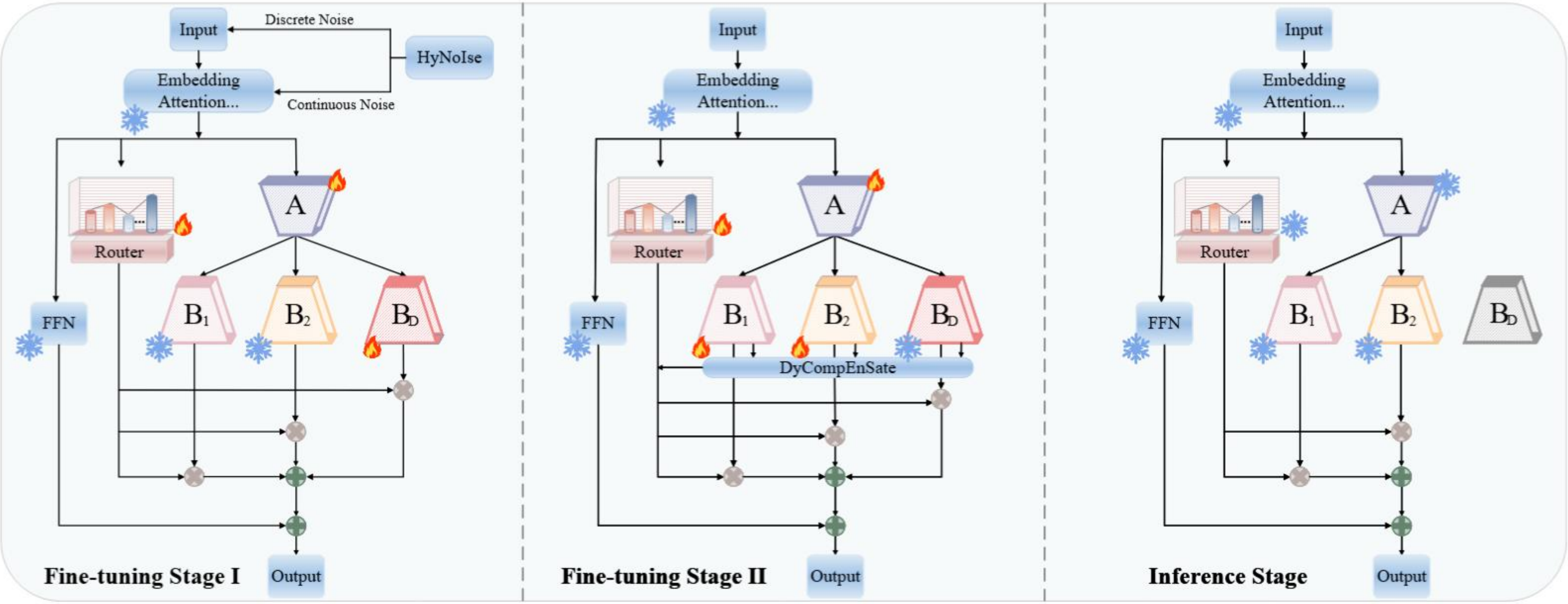
Mixture-of-Experts

and

HydraLoRA



Introduction



Main Figure

Introduction

Main contributions:

1. We propose LoPE, a novel noise-robust adaptation method that utilizes noise injection to handle noise.

2. We design a flexible hybrid noise injection strategy, introducing discrete noise at the input level and continuous noise at the embedding level.

3. Extensive experiments on multiple mainstream benchmark datasets.

Methodology

Hybrid Noise Injection (HyNoIse):

$S = \{(x_i, p_i)\}_{i=1}^M$

$\downarrow$

NoiseFunction}(·)

$\downarrow$

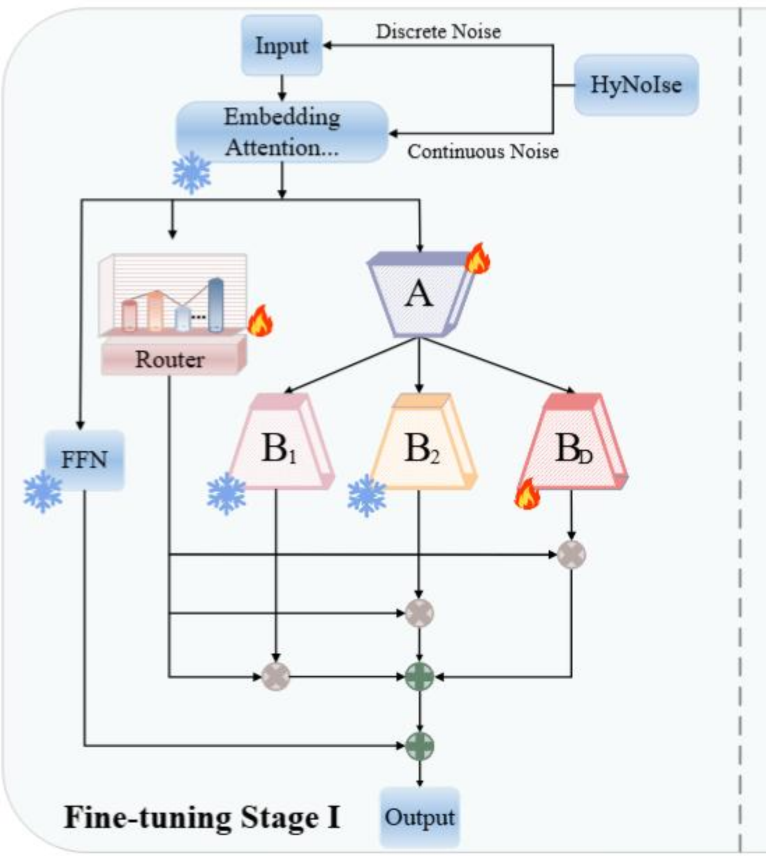
$S' = \{(x'_i, p'_i)\}_{i=1}^M$

discrete noise perturbations (word order shuffling, noise character insertion, and character deletion...)

continuous noise perturbations

Methodology

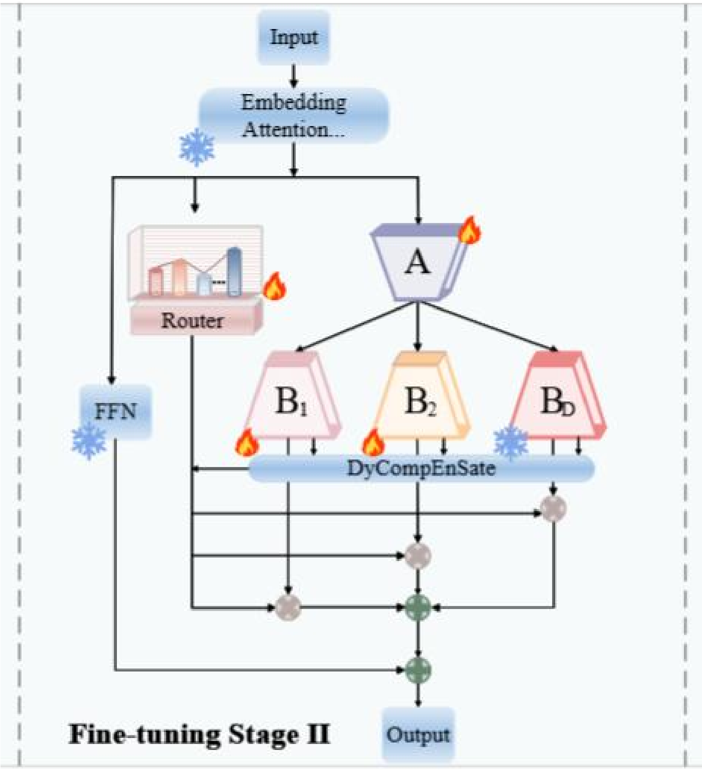
Fine-tuning Stage I: Specialized Poisoning Expert



$$y = W_0 x + (\omega_D B_D + \sum_{i=1}^{N-1} \omega_i f(B_i))Ax$$

Methodology

Fine-tuning Stage II: Dynamically Compensated Expert Synergy:

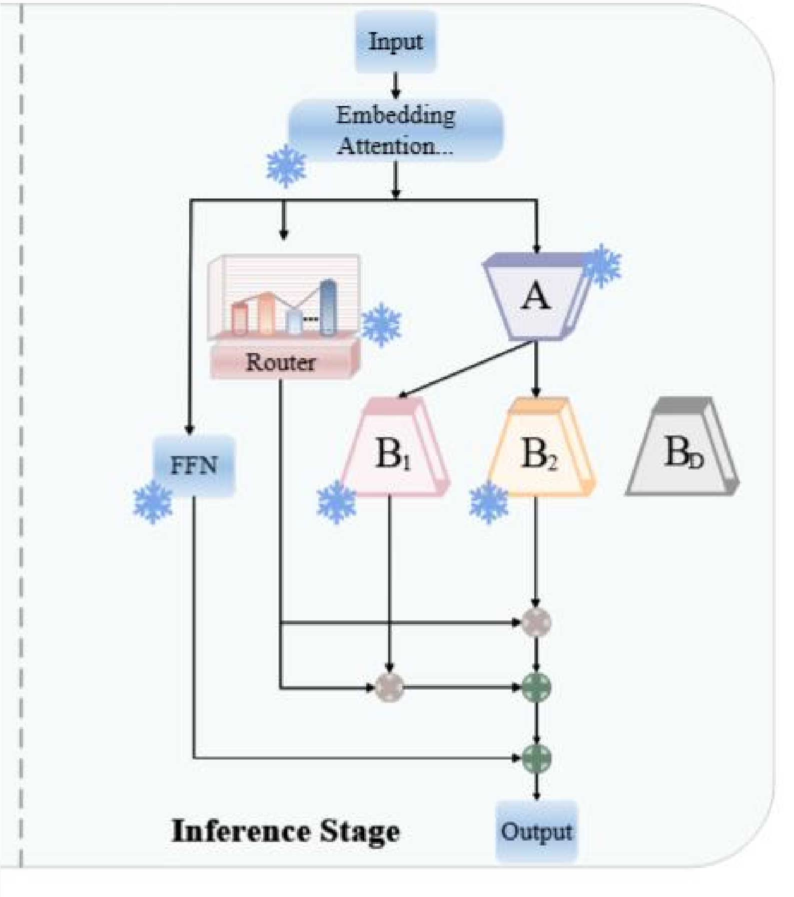


$$y = W_0 x + (\omega_D f(B_D) + \sum_{i=1}^{N-1} \omega_i B_i) A x$$

$$y = W_0 x + \beta(\omega_D f(B_D) + \sum_{i=1}^{K} (1 + \theta_{iD}) \omega_i B_i) A x$$

Methodology

Inference Stage:



$$y = W_0 x + \beta \sum_{i=1}^{K} (1 + \theta_{iD}) \omega_i f(B_i) A x$$

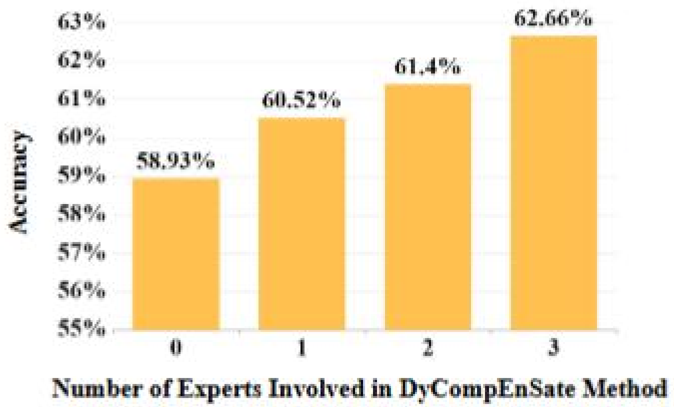
Experiments

Methods	MMLU	PIQA	SIQA	GSM8K	ARC-e	NSET	SBH	History	%Param	#A	#B
HydraLoRA(r=4)	46.10	76.28	52.92	16.15	62.61	35.99	**54.78**	55.95	0.062	1	3
LoPE(r=4)	**46.86**	**76.99**	**53.58**	**17.89**	**64.20**	**36.82**	54.46	**56.66**	0.062	1	3
P-Tuning†	37.23	71.65	39.97	8.87	45.21	26.04	37.14	45.09	0.193	-	-
Prefix Tuning†	37.91	71.79	40.42	9.25	43.48	27.13	38.77	44.21	0.077	-	-
AdaLoRA(r=2)†	39.11	72.29	41.07	10.38	47.84	29.16	39.83	49.69	0.023	1	1
LoRA(r=2)†	38.22	69.47	40.94	9.13	46.24	26.17	38.61	46.83	0.015	1	1
LoRA(r=4)†	40.45	71.45	43.17	11.02	48.39	27.96	40.88	49.03	0.031	1	1
HydraLoRA(r=2)†	42.47	74.65	46.38	10.74	56.08	31.26	42.37	52.65	0.031	1	3
HydraLoRA(r=4)†	43.08	74.92	47.29	11.83	56.26	34.74	52.35	55.80	0.062	1	3
LoPE(r=2)†	$43.05_{\pm0.28}$	$75.03_{\pm0.30}$	$46.76_{\pm0.17}$	$11.23_{\pm0.37}$	$58.93_{\pm0.51}$	$33.36_{\pm0.47}$	$48.65_{\pm0.32}$	$54.72_{\pm0.29}$	0.031	1	3
LoPE(r=4)†	$43.76_{\pm0.20}$	$75.49_{\pm0.41}$	$48.33_{\pm0.30}$	$12.72_{\pm0.33}$	$58.66_{\pm0.46}$	$34.06_{\pm0.11}$	$48.98_{\pm0.37}$	$55.46_{\pm0.18}$	0.047	1	2
LoPE(r=4)†	$\mathbf{44.42}_{\pm0.18}$	$\mathbf{76.28}_{\pm0.38}$	$\mathbf{49.03}_{\pm0.41}$	$\mathbf{13.72}_{\pm0.34}$	$\mathbf{60.49}_{\pm0.27}$	$\mathbf{35.45}_{\pm0.33}$	$\mathbf{52.88}_{\pm0.20}$	$\mathbf{56.84}_{\pm0.46}$	0.062	1	3

Ablation Study (HyNoIse,Backbone,DyCompEnSate)

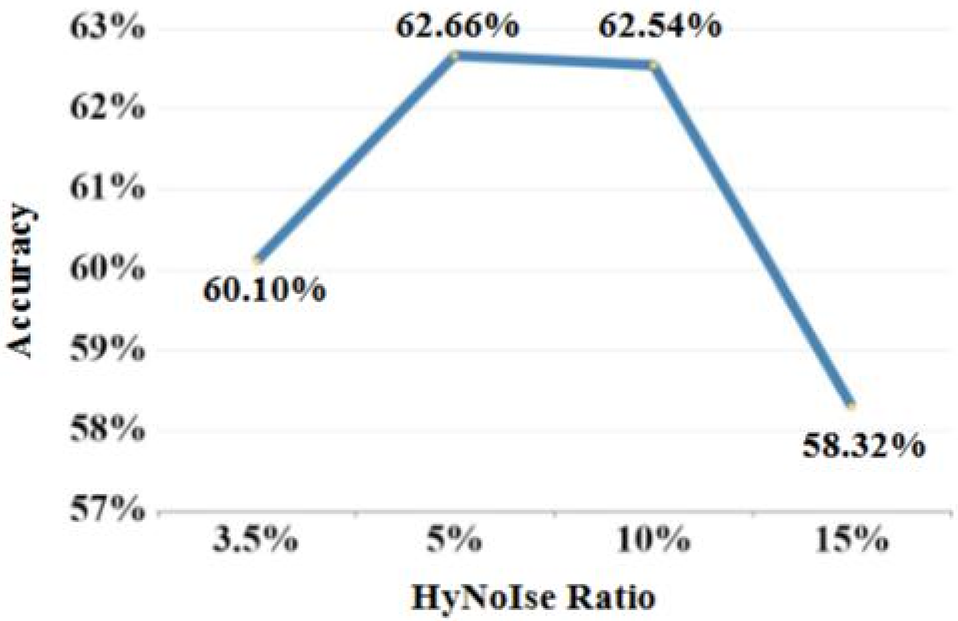
Noise Type	None	Continuous	Discrete	Hybrid
3.5% Level	60.90	61.23	62.07	**63.31**
5% Level	59.89	60.71	61.95	**62.66**
8% Level	57.48	58.68	60.14	**61.86**

Approaches	T5-large	LLaMA2-7b	Qwen2-7b	Qwen1.5-14b
HydraLoRA	33.64	47.29	66.49	78.23
LoPE	**36.37**	**49.03**	**68.99**	**80.02**



Analysis

Does Higher HyNoise Ratio Enhance Performance?



Analysis

Can the Poisoning Expert Truly Accomplish Its Task?

Method	PE=3,NE=1	PE=3,NE=2	PE=2,NE=2	Mask(PE=3,NE=1)	Not Mask(PE=3,NE=1)
LoPE	**62.66**	62.31	60.31	**62.66**	60.75

NEURAL INFORMATION PROCESSING SYSTEMS

Thank You!