NEURAL INFORMATION PROCESSING SYSTEMS
2025

Learning to Generalize: An Information Perspective on Neural Processes

Hui Li

State Key Laboratory of Advanced Rail Autonomous Operation, Beijing, China

School of Computer Science and Technology, Beijing Jiaotong University, Beijing, China

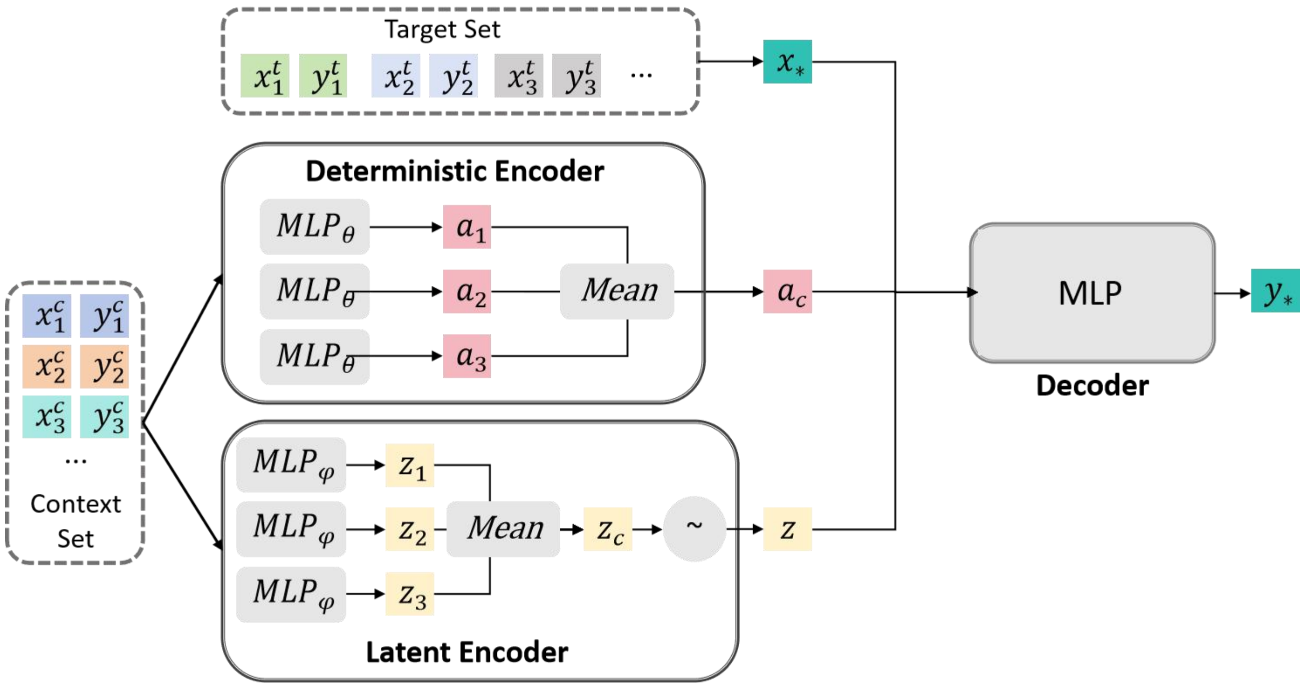
Beijing Key Laboratory of Traffic Data Mining and Embodied Intelligence, Beijing, China

Joint work with Huafeng Liu, Shuyang Lin, Jingyue Shi, Yiran Fu, and Liping Jing

Neural Processes

◆What is Neural Processes

- ***Probabilistic meta-learning framework:*** Neural Processes are **probabilistic models** designed to meta-learn distributions over functions, offering a data-driven alternative to **Gaussian Processes**.
- ***Meta-learning paradigm:*** Each function prediction is treated as a task, where models learn from context points during training and **generalize to new functions** with minimal observations at test time.



Neural Processes Generalization

- What is Neural Processes generalization challenge
  - ***Empirical-focused development***: Current NPs achieve strong performance through architectural innovations but **lack rigorous theoretical understanding** of generalization capabilities.
  - ***Theoretical gap***: Existing methods prioritize **performance optimization** while **neglecting theoretical interpretability** of generalization across diverse tasks.

- Previous solutions (architectural & meta learning):
  - Enhance model expressiveness through advanced mechanisms

    ANPs, TNPs, ConvCNPs, NDPs, NPCL
  - Improve learning efficiency with limited theoretical foundation

    Focus on inductive biases and model complexity without generalization theory

Our work focuses on providing a rigorous **information-theoretic framework** for understanding and improving NPs generalization via **dynamical stability regularization** and **optimization-aware noise injection**

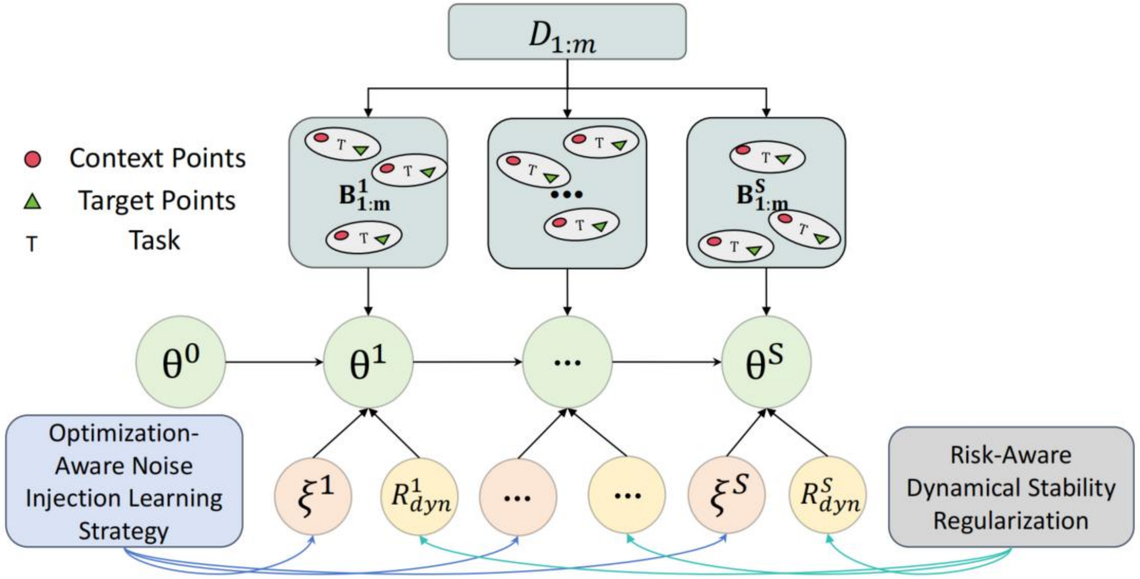
Information-Theoretic Generalization Method

◆ Information-Theoretic Bound

- ***Theorem 4.1***: If loss $\ell(\theta, X \times Y)$ is σ-subgaussian, then $\left|gen_{meta}^{NPs}(\tau, \mathcal{A})\right| \leq \sqrt{\frac{2\sigma^2}{mn} I(\theta; \mathcal{D}_{1:m})}.$
- ***Key Insight***: Minimizing mutual information $I(\theta;\ D_{1:m})$ improves generalization

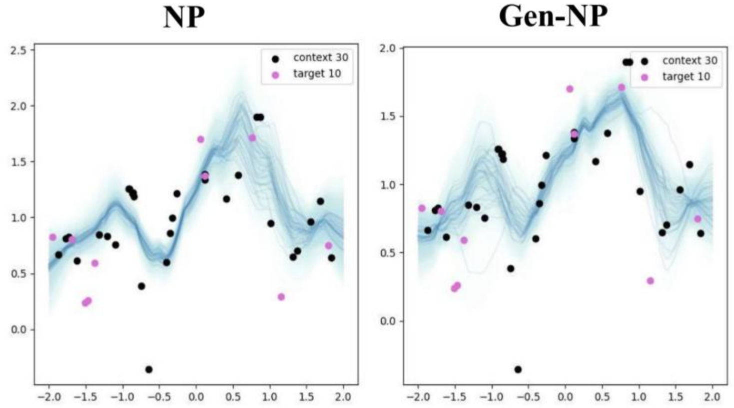
◆ Combined Effects

- ***Enhanced Stability***: $R_{\text{dyn, noise}} \geq \lambda_1 \cdot \mathbb{E}[Tr(H)] \cdot (1 + \eta_s \sigma_s^2) + \lambda_2 \cdot \mathbb{E}[\|H\|_F] \cdot \sqrt{1 + \eta_s \sigma_s^2}$
- ***Reduced MI***: $I_{\text{noise}}(\theta; \mathcal{D}_{1:m}) \leq \max\left(0, I(\theta; \mathcal{D}_{1:m}) - \eta_s \sigma_s^2 \cdot \mathbb{E}[\|\nabla \tilde{R}_{\mathcal{D}_i^T}(\theta)\|^2]\right).$

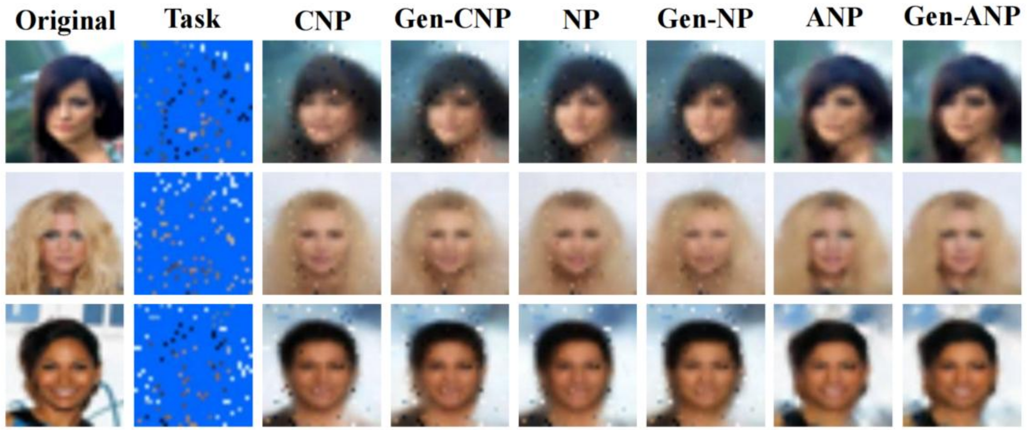


Result: Tighter generalization bound through joint optimization of stability and information

Experiments



(a) 1D Regression Results on NP



(b) Image Completion Results on CelebA Datasets

Table 3: Ablation study results comparing the original method.

Method	LL (RBF)	GI (RBF)	LL (Periodic)	GI (Periodic)
Original CNP	$0.265_{\pm 0.015}$	$0.880_{\pm 0.027}$	$-1.435_{\pm 0.020}$	$1.312_{\pm 0.053}$
Gen-CNP (with DSR only)	$0.276_{\pm 0.013}$	$0.858_{\pm 0.030}$	$-1.428_{\pm 0.022}$	$1.202_{\pm 0.045}$
Gen-CNP (with NILS only)	$0.279_{\pm 0.012}$	$0.846_{\pm 0.035}$	$-1.423_{\pm 0.024}$	$1.151_{\pm 0.041}$
Full Gen-CNP (DSR + NILS)	$\mathbf{0.286_{\pm 0.010}}$	$\mathbf{0.830_{\pm 0.042}}$	$\mathbf{-1.418_{\pm 0.023}}$	$\mathbf{1.112_{\pm 0.039}}$

- Classical NPs Benchmark: 1-D Regression, Image Completion, Bayesian Optimization, Contextual Bandits
- Ablation Study
- Comparison with Stability Neural Processes

Thanks