Neural Laplace Conformal Prediction (Research Proposal)

Bofan Chen
Department of Pure Mathematics and Mathematical Statistics
University of Cambridge
bc600@cam.ac.uk

1 Our goal

The goal of this research is to predict uncertainty on time series.

Notation and assumption For a system with $D \in \mathbb{N}^+$dimensions, the state of dimension d at time t is denoted as $y_d(t), \forall d = 1, \ldots, D, \forall t \in \mathbb{R}$. We elaborate that the trajectory $y_d : \mathbb{R} \to \mathbb{R}$ is a function of time. We consider the state vector at any time t, which is $\mathbf{y}(t) := [y_1(t), \ldots, y_D(t)]^T \in \mathbb{R}^D$. We further assume that the trajectory follows DE dynamics

Data. For each trajectory $\mathbf{y}^{(i)}(t)$, the state observations are made at discrete times of $t \in \mathcal{T} = \{t_1, t_2, \ldots, t_K, t_{K+1}, \ldots, t_{K+H}\}$; and there exists a certain time T, such that $t_1 < t_2 < \ldots < t_K < T < t_{K+1} < \ldots < t_{K+H}$. Our dataset should be expressed like this:

$$\mathcal{D}_{\text{labeled}} = \{\{(t_k^{(i)}, \mathbf{y}_k^{(i)})\}_{k=1}^{K+H}\}_{i=1}^{l}.$$

We can use T to divide the trajectory into the known part and the predicting part by letting

$$\mathcal{D}_{\text{labeled}} = \{X^{(i)}, P^{(i)}\}_{i=1}^{l},$$

where

$$X^{(i)} = \{(t_k^{(i)}, \mathbf{y}_k^{(i)})\}_{k=1}^{K}$$

and

$$P^{(i)} = \{(t_k^{(i)}, \mathbf{y}_k^{(i)})\}_{k=K+1}^{K+H}.$$

Input and Output. We design a neural network architecture to use the trajectory observations X before time T to predict the trajectory $\hat{\mathbf{y}}(t, X), \forall t > T$. Furthermore, we also estimate the uncertainty of the trajectory which can be expressed by the confidence interval

$$\hat{C}^{\alpha}(t, X) = [\hat{\mathbf{y}}(t, X) - \epsilon\hat{\sigma}(t, X), \hat{\mathbf{y}}(t, X) + \epsilon\hat{\sigma}(t, X)],$$

where

$$\mathbb{P}(\mathbf{y}(t) \in \hat{C}^{\alpha}(t, X) | \mathbf{y}(t_k) = \mathbf{y}_k, \forall(t_k, \mathbf{y}_k) \in X) > 1 - \alpha.$$

2 The model

Divide the dataset.

$$\mathcal{D}_{\text{labeled}} = \{\mathcal{D}_{\text{train}} \cup \mathcal{D}_{\text{res}} \cup \mathcal{D}_{\text{cal}} \cup \mathcal{D}_{\text{test}}\}$$

Encode the input. Obtain the initial state of each trajectory:

$$\mathbf{p} = h_{\gamma}(X),$$

37th Conference on Neural Information Processing Systems (NeurIPS 2023).

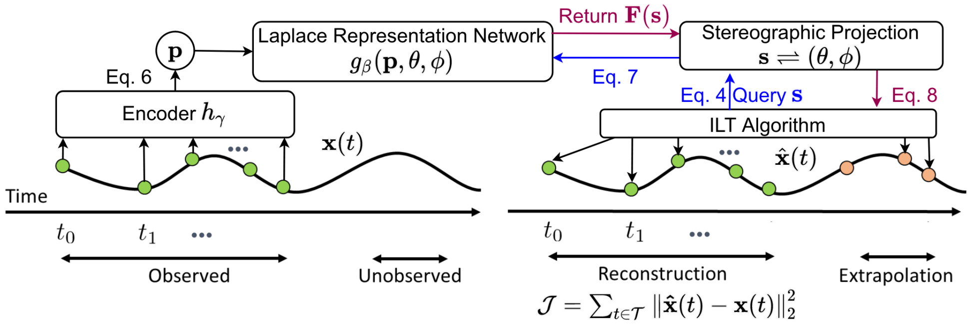

Figure 1: Neural Laplace.

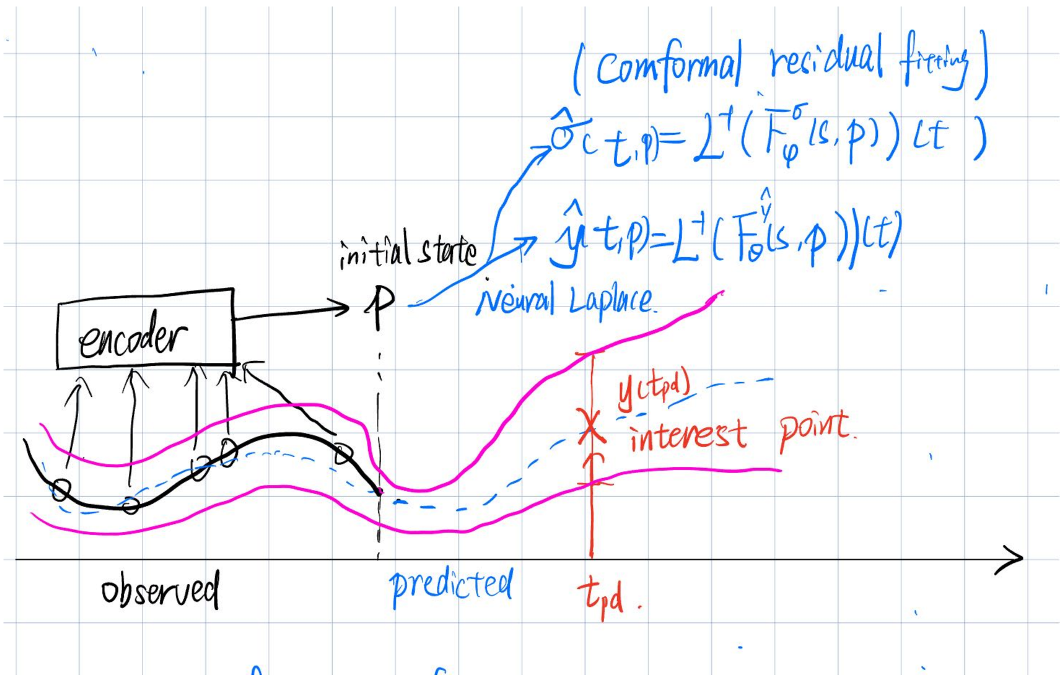

Figure 2: Neural Laplace Conformal Prediction.

where $h_\gamma(\cdot)$ can use RNN, LSTM, GRU.

Train the predictive model. Use $\mathcal{D}_{\text{train}}$ and Neural Laplace method to obtain the predict trajectory $\hat{\mathbf{y}}(t, \mathbf{p})$.

Train the conformal normalization model. Use $\mathcal{D}_{\text{res}}$ and Neural Laplace method to obtain the conformal normalization model $\hat{\sigma}(t, \mathbf{p})$.

Conformal prediction. Use $\mathcal{D}_{\text{cal}}$ to obtain the confidence interval.

Non-conformity function:

$$\gamma(\mathbf{p}) = \sup_{k \in 1:K+H} \left| \frac{\mathbf{y}_k - \hat{\mathbf{y}}(t_k, \mathbf{p})}{\hat{\sigma}(t_k, \mathbf{p})} \right|,$$

where $\mathbf{p} = h_\gamma(X), X \in \mathcal{D}_{\text{cal}}^X$.

Non-conformity score ϵ: the $\lceil (|\mathcal{D}_{\text{cal}}| + 1)(1-\alpha) \rceil$-th smallest in $\gamma(\mathbf{p})$.

Confidence interval for the initial state $\mathbf{p}$:

$$\hat{C}^\alpha(t, X) = [\hat{\mathbf{y}}(t, \mathbf{p}) - \epsilon \hat{\sigma}(t, \mathbf{p}), \hat{\mathbf{y}}(t, \mathbf{p}) + \epsilon \hat{\sigma}(t, \mathbf{p})],$$

where $\mathbf{p} = h_\gamma(X)$.

3 Advantages

Satisfy exchangeability. Assumption to use conformal prediction: *(Exchangeability) In a dataset of l observations $\left\{\left(\mathbf{x}^{(i)}, y^{(i)}\right)\right\}_{i=1}^{l}$, any of its l ! permutations are equiprobable. Note that independent identically distributed (iid) observations satisfy exchangeability.* For the time series dataset, $(t_k^{(i)}, \mathbf{y}_k^{(i)})$ in $X^{(i)}$ does not satisfy the exchangeability, because the y_k may be related in one single time series. We treat each time series as one observation which satisfy the exchangeability.

Locally adaptive. For different initial state, the uncertainty in the future is different. We set the score function wisely to ensure it is locally adaptive.

Irregular time step. Our framework use Neural Laplace control method which can be use in the irregular time dataset. Furthermore, when we calculate the confidence interval, we do not need that the predicting part of the labeled data is observed at regular interval.

Computational conplexity. Our framework use Neural Laplace control method which modeled the function in whole time line. This greatly reduces the computational cost.

4 Experiments

1. For simplicity, consider the time series that $d = 1$ and regular time step.

5 Futher research

- consider observation have noise.
- consider self-supervised conformal prediction.
- consider the MPC controller over the modeln where setting the reward function related to uncertainty.