Rough Transformers for Continuous Time Series and Hurst index Classification

Eliott Pradeleix Louis Perot

Poster made as part of a work for the MVA course Deep Learning (V.Lepetit, M. Vakalopoulou)

Objectives

1. Design a Transformer architecture that operates on continuous-time representations via path signatures.
2. Develop a multi-view attention mechanism that jointly encodes global and local temporal structure in long time series.
3. Evaluate the proposed architecture on long and irregularly sampled time series, with a focus on accuracy, robustness, and computational efficiency.

Introduction

Real-world sequential data are often long, irregularly sampled, and of variable length. Continuous-time models and Transformer-based architectures address these challenges by modelling irregular sampling and long-range dependencies, respectively. However, Neural ODE-based models [1] [2] are computationally expensive for long sequences, while Transformers rely on discrete-time representations and require fixed length. This work [3] extends the Transformer architecture using the signature transform from *Rough Path theory*, achieving improved representational efficiency and performance with reduced computational cost.

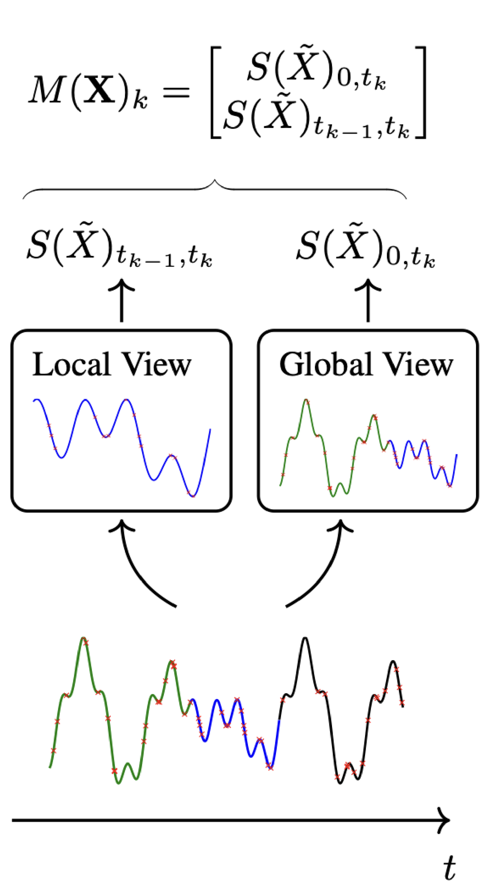

Figure 1:Multi-view signature transform.

Background

Problem formulation. $X : [0,T] \to \mathbb{R}^d$ be a continuous-time process observed at irregular times $\{(t_i, X(t_i))\}_{i=1}^L$. We aim to learn a functional $f(X)$ for a downstream task, without assuming fixed L.

The Transformer architecture [4]. Given $X \in \mathbb{R}^{L\times d}$, a linear positional encoding $Z = T(X)$ is applied, followed by multi-head self-attention

$$\mathrm{Attn}(Z) = \mathrm{softmax}\left(\frac{ZW^Q(ZW^K)^\top}{\sqrt{d_k}}\right) ZW^V,$$

and a feed-forward output projection.

Rough Path Signatures. For a path $X : [0,T] \to \mathbb{R}^d$ and a word $\mathcal{I} = (i_1, \dots, i_k)$, define

$$S^{\mathcal{I}}(X)_{s,t} = \int_{s<u_1<\dots<u_k<t} dX^{i_1}_{u_1} \cdots dX^{i_k}_{u_k}.$$

The signature transform is $S(X)_{s,t} = \{S^{\mathcal{I}}(X)_{s,t}\}_I$.

Experiments

Hurst index classification. Fractional Brownian motion (fBm) is a self-similar Gaussian process parametrized by the Hurst index $H \in (0,1)$, with $H = \frac{1}{2}$ corresponding to standard Brownian motion. Estimating H from discrete observations is challenging due to non-stationarity and long-range dependence [5]. We train models to learn the mapping $X = \left((t_0, B^H_{t_0}), \dots, (t_n, B^H_{t_n})\right) \mapsto f(X) = H$ in a classification setting.

model	accuracy (%)	training time (s)
LSTM	10.00	174.59
NCDE	19.33	60.78
Transformer	24.00	71.72
RFormer-G	40.67	22.53
RFormer-GL	53.33	22.34
RFormer-L	**58.67**	**22.66**

Important Result

Let $\mathbb{T}$ be a Rough Transformer and let $\hat{X} : [0,T] \to \mathbb{R}^d$ be a continuous-time path. Let $\gamma : [0,T] \to [0,T]$ be a time reparameterization. If X and X' are (possibly irregular) samplings of $\hat{X}$ and $\hat{X} \circ \gamma$, respectively, then $\mathbb{T}(X) \approx \mathbb{T}(X')$.

Rough Transformers

Given an irregularly sampled time series X, we form a continuous-time interpolation $\tilde{X}$ and define the multi-view signatures

$$M(X)_k = (S(\tilde{X})_{0,t_k},\ S(\tilde{X})_{t_{k-1},t_k}).$$

Truncated signatures replace discrete tokens in a Transformer, with attention given by

$$\mathrm{Attn}(M) = \mathrm{softmax}\left(\frac{(MW^Q)(MW^K)^\top}{\sqrt{d_k}}\right) MW^V.$$

We call Rough Transformer [3], and denote $\mathbb{T}$, the Transformer architecture with this new attention mechanism.

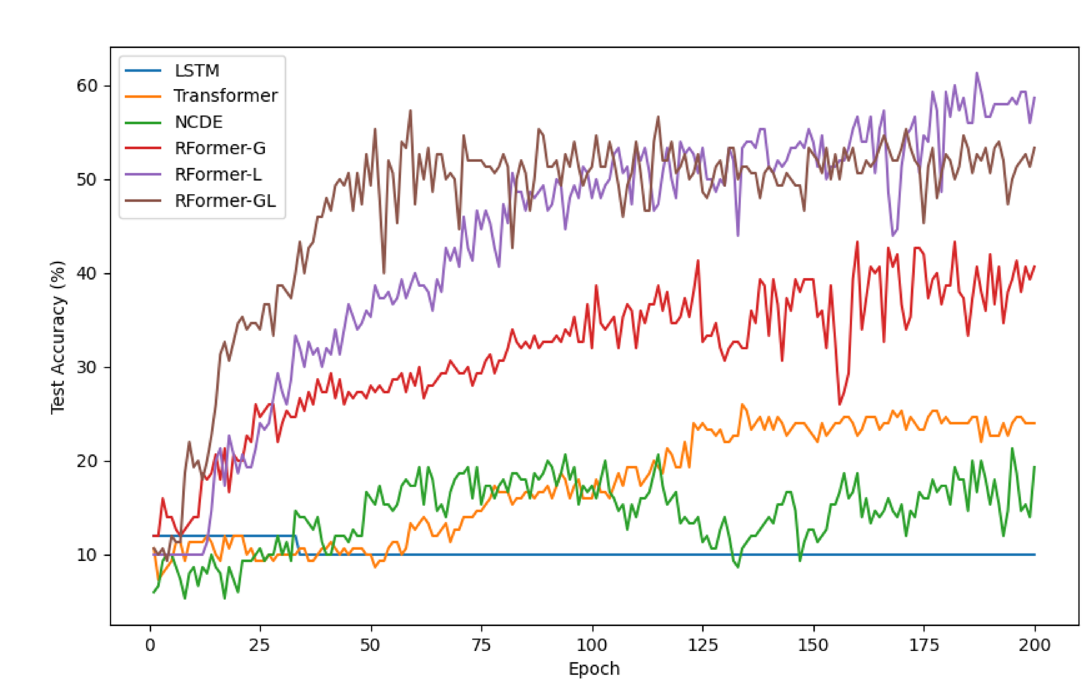

Figure 2:Test accuracy vs. epoch during the training phase.

Results

RFormer consistently outperforms the baselines in both test accuracy and training efficiency. Moreover, incorporating the local-view mechanism further improves performance.

Irregular time series classification. A key property of RFormer is its ability to naturally handle irregularly sampled time series due to time-reparameterization invariance, which suggests robustness to subsampling.

	accuracy (%)	
model	50% drop	70% drop
LSTM	10.00	10.00
NCDE	15.33	12.67
Transformer	20.00	16.67
RFormer-G	34.00	30.00
RFormer-GL	42.67	24.67
RFormer-L	**51.33**	**42.00**

Conclusion

RFormer provides an efficient and robust approach to continuous-time time series modelling. By exploiting rough path signatures and time-reparameterization invariance, it achieves higher accuracy, faster training, and strong resilience to subsampling compared to standard baselines.

References

[1] Kidger et. al : Neural controlled differential equations for irregular time series (2020).

[2] Morrill et al. : Neural rough differential equations for long time series (2021).

[3] Moreno-pino et. al : Rough transformers: Lightweight and continuous time series modelling through signature patching (2024).

[4] Vaswani et al. : Attention is all you need (2017).

[5] Bonnier et al. : Deep signature transforms (2019).