Towards Understanding Evolving Patterns in Sequential Data

Qiuhao Zeng[1], Long-Kai Huang[2], Qi Chen[3], Charles X. Ling[1], Boyu Wang[1]

[1] University of Western Ontario [2] Tencent AI Lab [3] Laval University

Motivation

- **How can the existence of evolving patterns in data sequences be determined?**

  Consider the scenario of a person repeatedly tossing a coin. In this case, historical information does not influence the outcome of the next toss.

- **Can one determine the historical span that significantly influences the current time point?**

  How do we determine the *order*, i.e., the optimal number of past observations, of an autoregressive model in a principled way?

- **How can we determine if the collected features are sufficient to reveal evolving patterns?**

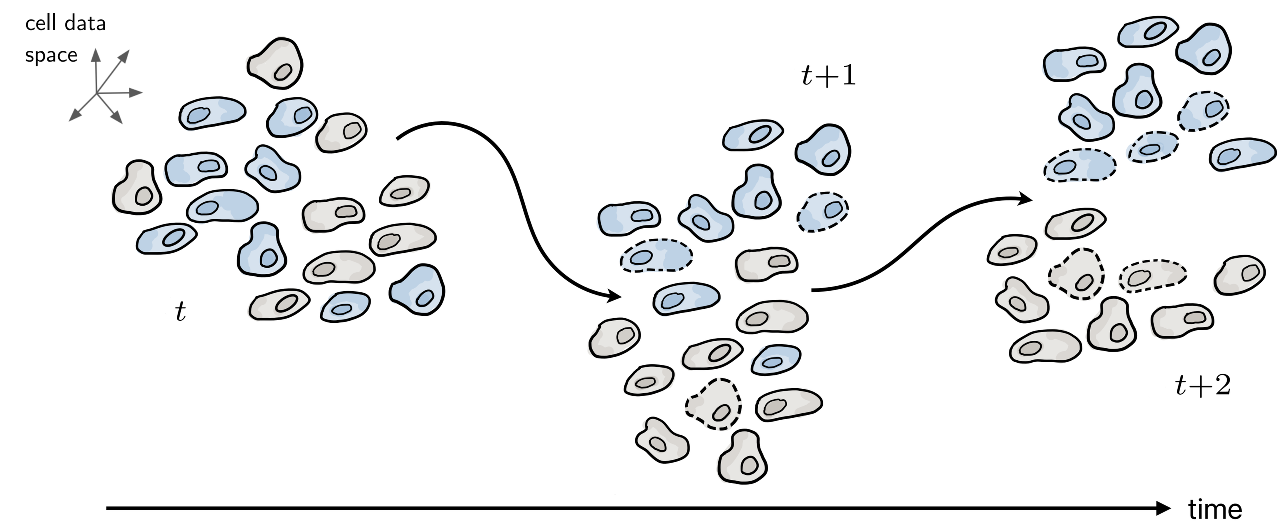
  For instance, to achieve better weather forecasting, how can one determine the essential features, such as altitude, humidity, and geographic location, for gathering a comprehensive set of information for forecasting?

Contributions

- We propose **EvoRate**, which enables quantitatively measuring the evolving patterns existing in high-dimensional sequential data by utilizing the *neural mutual information estimator*.
- **EvoRate** can be applied to assess temporal order and conduct feature selections in sequential data.
- We further proposed **EvoRate**$_{\mathcal{W}}$ to leverage optimal transport to build the correspondence between snapshots at the different timestamps, and hence allow the Mutual Information (MI) approximations.

Motivation II

- The figure[a] illustrates that, without tracking cell trajectories, we only observe clusters of data at different timestamps, making it essential to estimate **correspondence** between time points to uncover individual cells' evolving patterns.



[a] Source: Charlotte et al., Optimal transport in learning, control, and dynamical systems. ICML Tutorial 2023.

EvoRate measures evolving patterns via MI

- **EvoRate** estimates the empirical sequential MI $\hat{I}(\mathbf{Z}_{t-k+1}^{t}; Z_{t+1})$ by defining $m : \mathbb{R}^{k\times D} \times \mathbb{R}^{D} \to \mathbb{R}$, $m(x_1^k, y) = -\|f(g(x_1), .., g(x_k)) - g(y)\|_2^2$:

$$\begin{aligned}\text{EvoRate} := &\ \mathbb{E}_{\mathbf{z}_{t-k+1}^{t+1}\sim P(\mathbf{Z}_{t-k+1},\ldots,Z_{t+1})} - \|f(g(z_{t-k+1}),\ldots,g(z_t)) - g(z_{t+1})\|_2^2 \\ &- \log \mathbb{E}_{\mathbf{z}_{t-k+1}^{t}\sim P(Z_{t-k+1},\ldots,Z_t), z_{t+1}\sim P(Z_{t+1})} e^{-\|f(g(z_{t-k+1}),\ldots,g(z_t)) - g(z_{t+1})\|_2^2}, \qquad (1)\end{aligned}$$

  where $g : \mathbb{R}^D \to \mathbb{R}^d$ is an encoder.
- **Proposition** Let H denote the entropy. For autoregression tasks, the expected MLE loss satisfy:

$$\mathcal{L}_{mle} = \underbrace{D_{\mathrm{KL}}(P(Z_{t+1}|\mathbf{Z}_{t-k+1}^{t}), Q(Z_{t+1}|\mathbf{Z}_{t-k+1}^{t}))}_{\text{(i) Model related}} + \underbrace{H(Z_{t+1}) - I(Z_{t+1}; \mathbf{Z}_{t-k+1}^{t})}_{\text{(ii) Data related}}$$

Estimate the absent correspondences

- The distance loss according to a joint distribution measurement π

$$\mathcal{L}_{\mathcal{W}}^{t}(\pi, f) = \mathbb{E}_{(z_t, z_{t+1})\sim\pi}\|f(g(z_t)) - g(z_{t+1})\|_2^2 \qquad (2)$$

  where g is fixed from updated gradients computed from $\mathcal{L}_{\mathcal{W}}^{t}$.
- The optimal transport plan π^* to approximate the real joint distribution

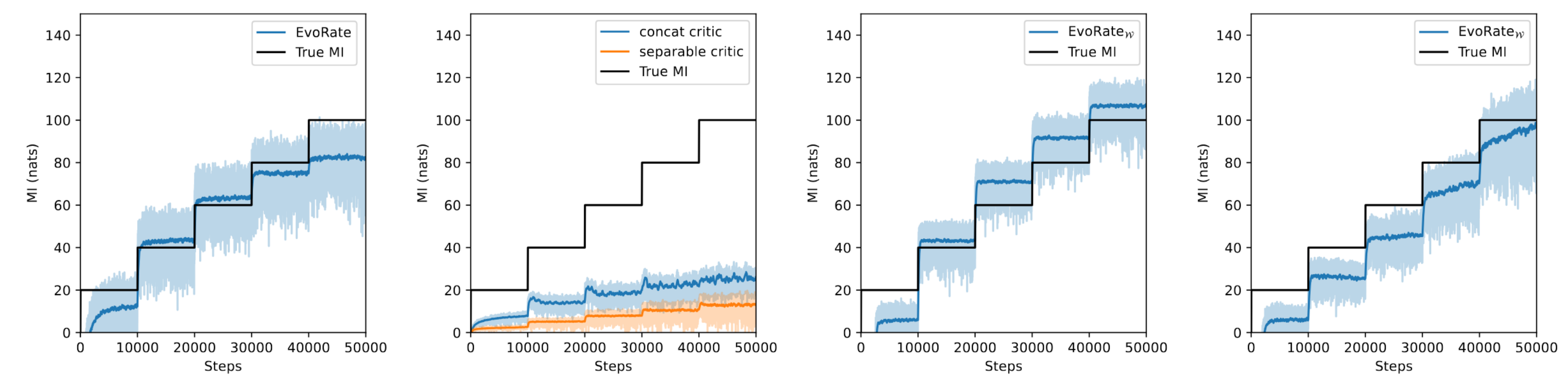
$$\pi^*(Z_t, Z_{t+1}) = \underset{\pi\in\Pi(P(Z_t), P(Z_{t+1}))}{\arg\min} \mathcal{L}_{\mathcal{W}}^{t}(\pi, f), \quad \forall t \in \{1, \ldots, T-1\}, \qquad (3)$$

EvoRate$_{\mathcal{W}}$ for w/o correspondence cases

- Use $\pi^*(Z_t, Z_{t+1})$ to estimate joint distribution P, and then obtain the following estimator with $\pi^*(Z_t, Z_{t+1})$

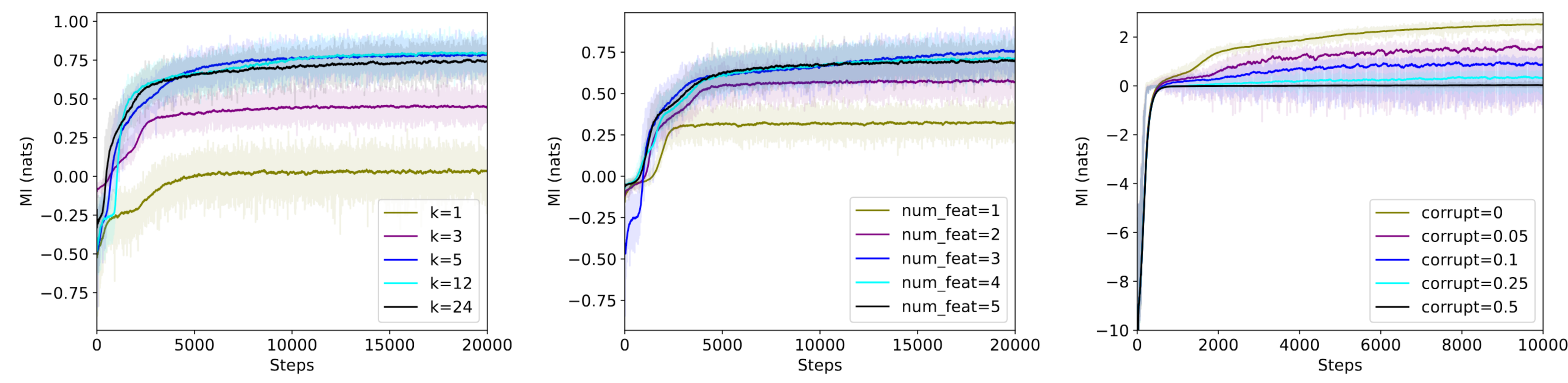
$$\begin{aligned}\text{EvoRate}_{\mathcal{W}} = &\ \mathbb{E}_{(z_t, z_{t+1})\sim\pi^*(Z_t, Z_{t+1})} - \|f(g(z_t))) - g(z_{t+1})\|_2^2 \\ &- \log \mathbb{E}_{z_t\sim P(Z_t), z_{t+1}\sim P(Z_{t+1})} e^{-\|f(g(z_t)) - g(z_{t+1})\|_2^2}\end{aligned}$$

Empirical Results



(a) (b) (c) (d)

- Figure 1. (a) k-order EvoRate estimation. (b) EvoRate estimation on a different number of features. (c) EvoRate estimation of the video prediction tasks with a different corruption rate.



(a) (b) (c)

- Figure 2. (a) k-order EvoRate estimation. (b) EvoRate estimation on a different number of features. (c) EvoRate estimation of the video prediction tasks with a different corruption rate.

	Crypto	Player Traj.	M4-Monthly	M4-Weekly	M4-Daily
RMSE/sMAPE	6.91	1.16	11.93	7.25	2.99
LEs	0.026	0.052	0.011	0.013	0.020
Trend	0.02	0.01	0.48	0.13	0.05
Seasonality	0.00%	0.00%	66.34%	0.00%	0.00%
EvoRate	2.80	4.67	1.58	2.25	2.26

- Table 1. In the above table, a larger EvoRate consistently indicates a smaller potential prediction error (RMSE/sMAPE) for the dataset.

	RGaussian	Circle	Sine	RMNIST	Portraits	Caltran	PowerSupply
Invariant (Acc:%)	47.5	51.3	63.2	39.0	85.4	64.1	70.8
Evolving (Acc:%)	97.7	73.8	71.4	46.4	89.1	70.6	75.7
Acc_{EVO} - Acc_{INV} (%)	50.2	22.5	8.2	7.4	3.7	6.5	4.9
EvoRate$_{\mathcal{W}}$	1.58	0.58	0.54	0.95	0.25	0.28	0.46

- Table 2. The estimated mutual information for the evolving domains for different datasets. The reported results are the average accuracy of the multiple target domains.