Generalizing to Evolving Domains with Latent Structure-Aware Sequential Autoencoder

Tiexin Qin   Shiqi Wang   Haoliang Li

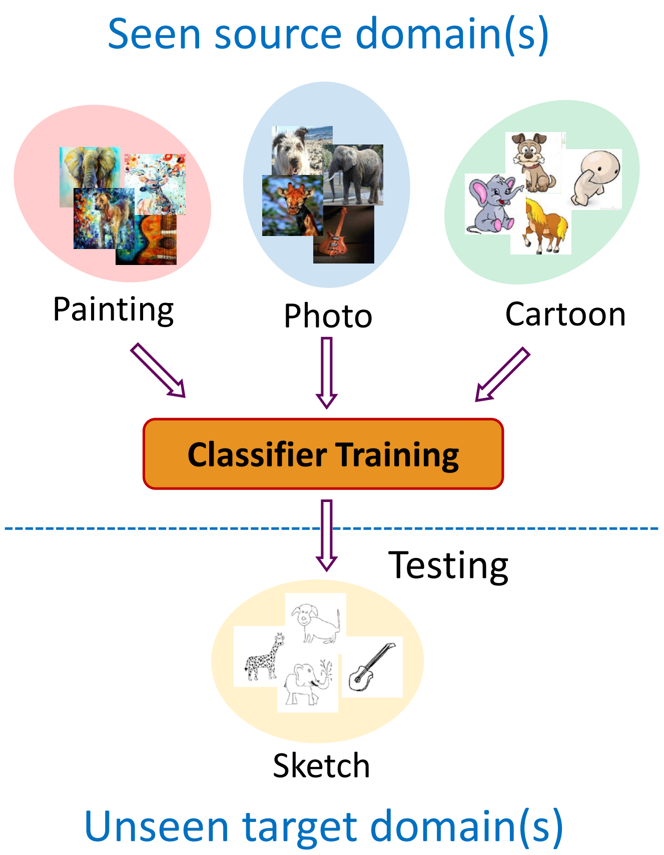
City University of Hong Kong

ICML 2022

ICML
International Conference
On Machine Learning

CityU
香港城市大學
City University of Hong Kong

Motivation: Domain Generalization



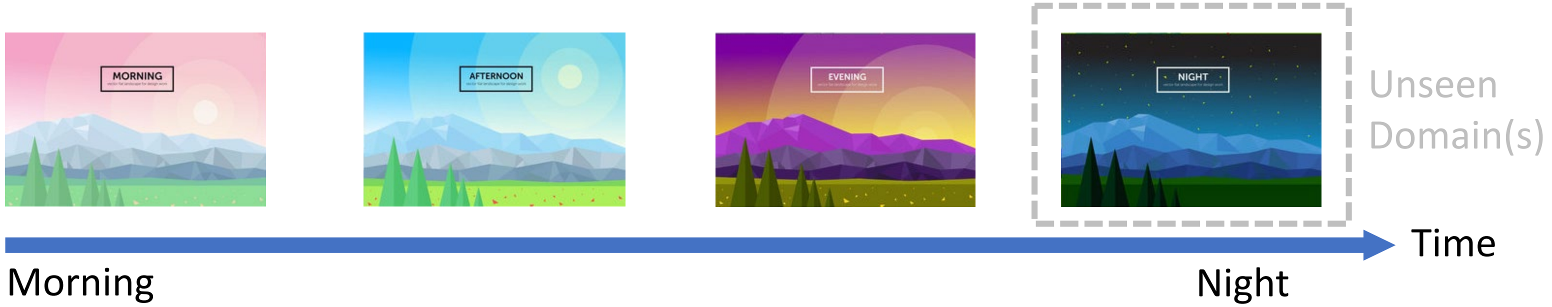
- Domain: a data generation regime
- **Domain generalization (DG)**: Build a robust recognition system for classification in previously unseen datasets, given one or multiple training datasets.

Requirement: No target samples available for model training

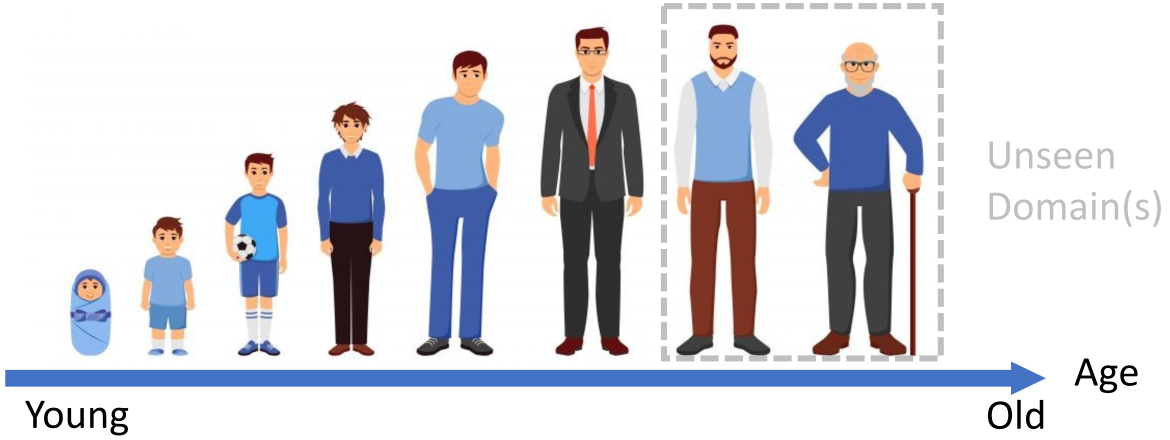
Weaknesses: static and discrete domains, inexplicit inter-domain correlation

Motivation: Evolving Domain Environment

- Self-driven car system

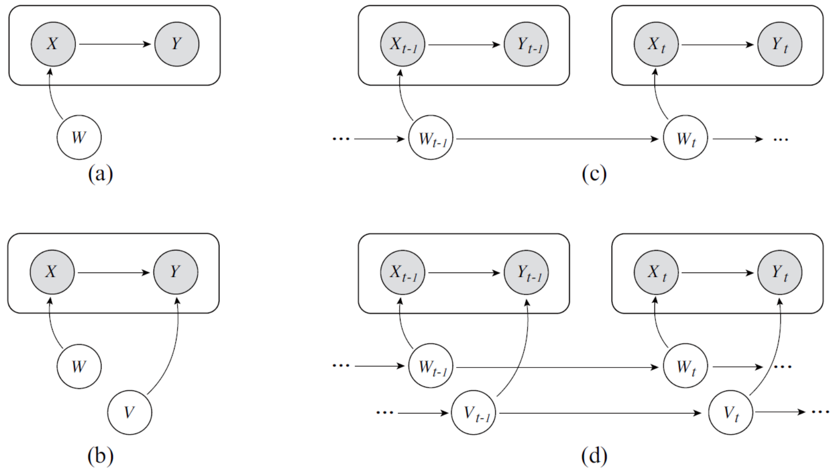


- Disease diagnosis



Methodology

- Latent Structure-aware Sequence AutoEncoder (LSSAE)



Bayesian rule:

$$P(X, Y) = P(X)P(Y|X)$$

Distribution shift:

(1) Covariate shift

$$P(X^s) \neq P(X^t)$$

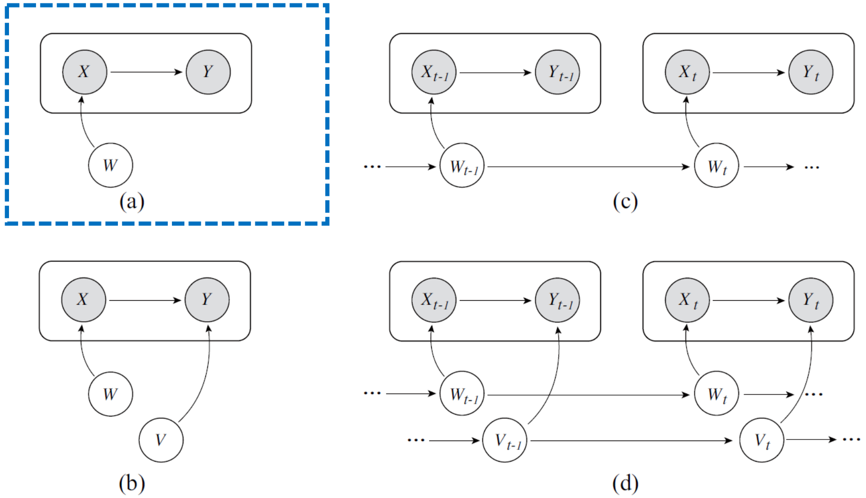
(2) Concept shift

$$P(Y^s|X^s) \neq P(Y^t|X^t)$$

Federici et al., "An Information-theoretic Approach to Distribution Shifts", NeurIPS'19.
Lu et al., "Learning under Concept Drift: A Review", TKDE'19.

Methodology

- Latent Structure-aware Sequence AutoEncoder (LSSAE)

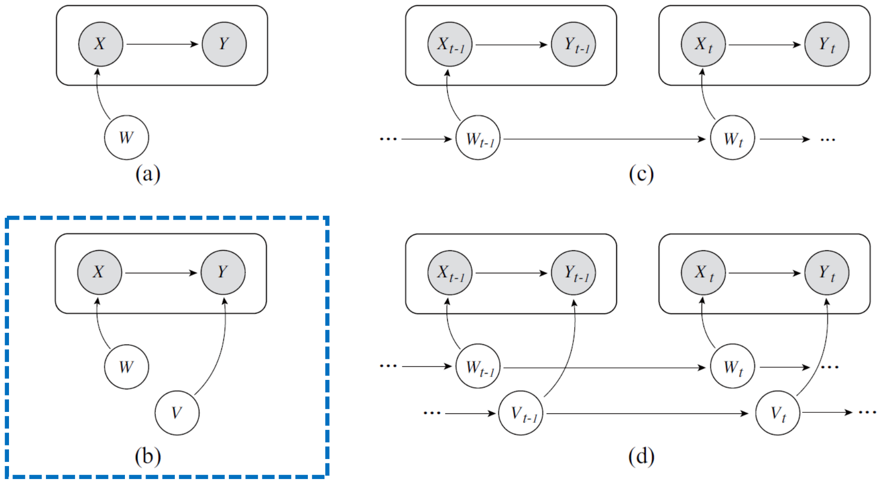


(a) Conventional DG with covariate shift only:

$P(X^s) \neq P(X^t)$ and $P(Y^s|X^s) = P(Y^t|X^t)$

Methodology

- Latent Structure-aware Sequence AutoEncoder (LSSAE)



(a) Conventional DG with covariate shift only:

$P(X^s) \neq P(X^t)$ and $P(Y^s|X^s) = P(Y^t|X^t)$

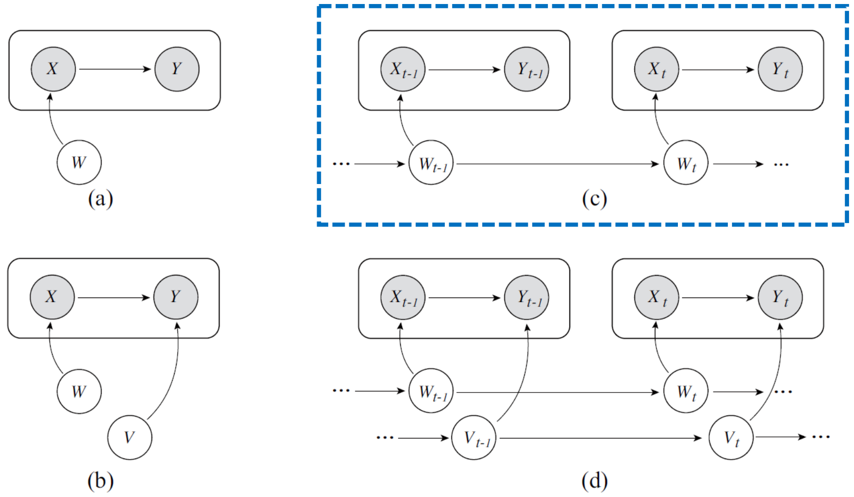
(b) Heterogenous DG:

$P(X^s) \neq P(X^t)$ and $P(Y^s|X^s) \neq P(Y^t|X^t)$

Li et al., “Feature-Critic Networks for Heterogeneous Domain Generalisation”, ICML’19.

Methodology

- Latent Structure-aware Sequence AutoEncoder (LSSAE)



(a) Conventional DG with covariate shift only:

$$P(X^s) \neq P(X^t) \text{ and } P(Y^s|X^s) = P(Y^t|X^t)$$

(b) Heterogenous DG:

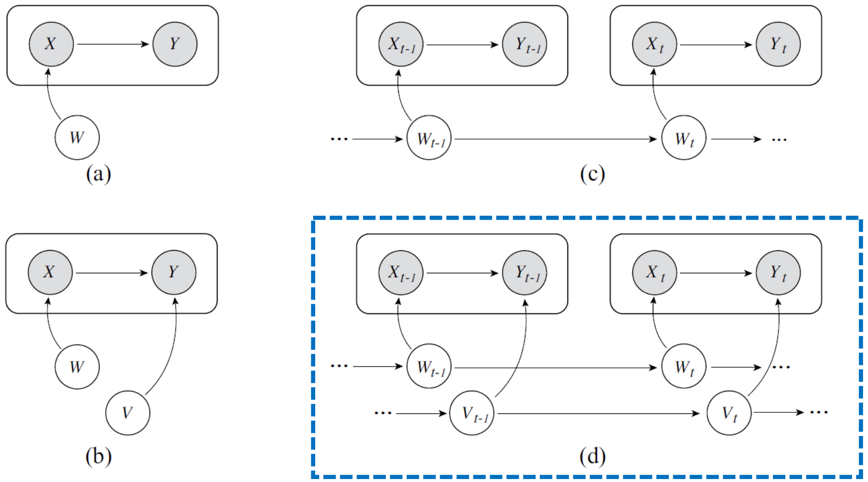
$$P(X^s) \neq P(X^t) \text{ and } P(Y^s|X^s) \neq P(Y^t|X^t)$$

(c) Evolving environment with covariate shift only:

$$P(X_{t-1}) \neq P(X_t) \text{ and } P(Y_{t-1}|X_{t-1}) = P(Y_t|X_t)$$

Methodology

- Latent Structure-aware Sequence AutoEncoder (LSSAE)



(a) Conventional DG with covariate shift only:

$P(X^s) \neq P(X^t)$ and $P(Y^s|X^s) = P(Y^t|X^t)$

(b) Heterogenous DG:

$P(X^s) \neq P(X^t)$ and $P(Y^s|X^s) \neq P(Y^t|X^t)$

(c) Evolving environment with covariate shift only:

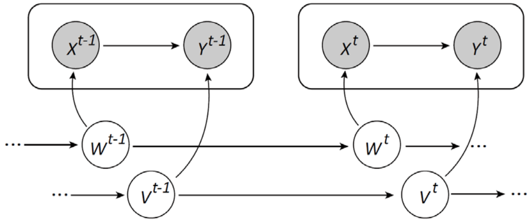
$P(X_{t-1}) \neq P(X_t)$ and $P(Y_{t-1}|X_{t-1}) = P(Y_t|X_t)$

(d) Evolving domain generalization (our focus):

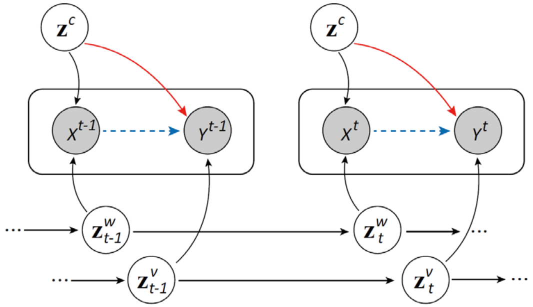
$P(X_{t-1}) \neq P(X_t)$ and $P(Y_{t-1}|X_{t-1}) \neq P(Y_t|X_t)$

Methodology

- Latent Structure-aware Sequence AutoEncoder (LSSAE)



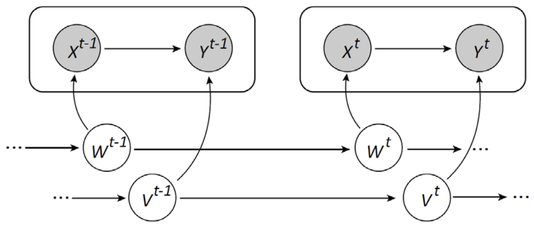
(d)



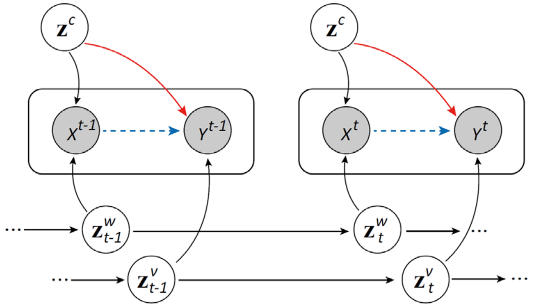
Probabilisitic Generative Model of (d)

Methodology

- Latent Structure-aware Sequence AutoEncoder (LSSAE)



(d)

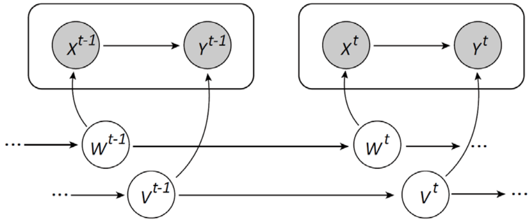


Probabilisitic Generative Model of (d)

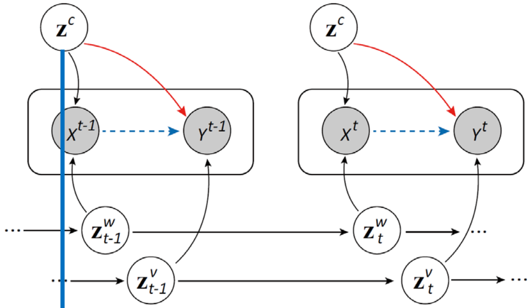
Defining latent codes →

Methodology

- Latent Structure-aware Sequence AutoEncoder (LSSAE)



(d)



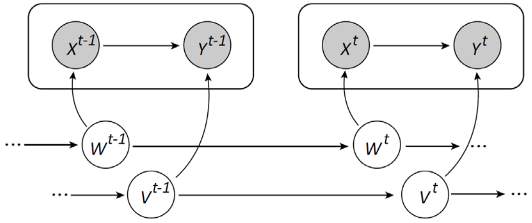
Probabilisitic Generative Model of (d)

$\boldsymbol{z}^C$: static domain-invariant category information from X

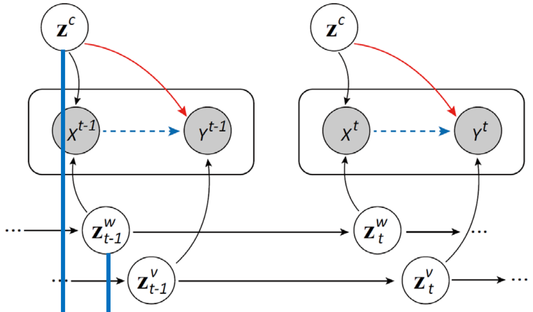
Defining latent codes

Methodology

- Latent Structure-aware Sequence AutoEncoder (LSSAE)



(d)



Probabilisitic Generative Model of (d)

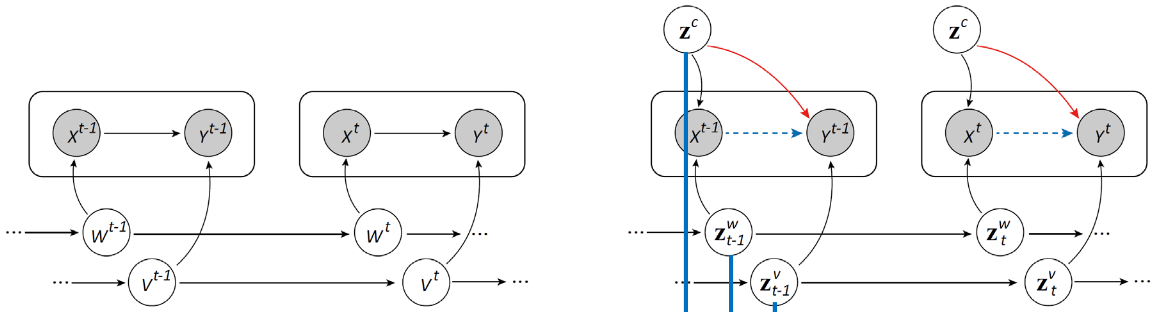
$\boldsymbol{z}^c$: static domain-invariant category information from X

Defining latent codes

$\boldsymbol{z}^w_{1:T}$: dynamic domain-specific information from X

Methodology

- Latent Structure-aware Sequence AutoEncoder (LSSAE)



(d)

Probabilisitic Generative Model of (d)

Defining latent codes

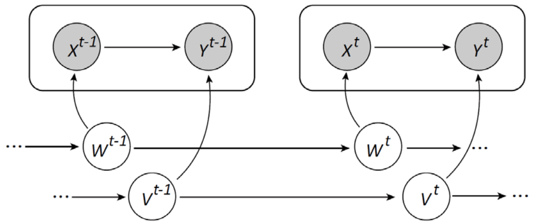
$\boldsymbol{z}^{c}$: static domain-invariant category information from X

$\boldsymbol{z}^{w}_{1:T}$: dynamic domain-specific information from X

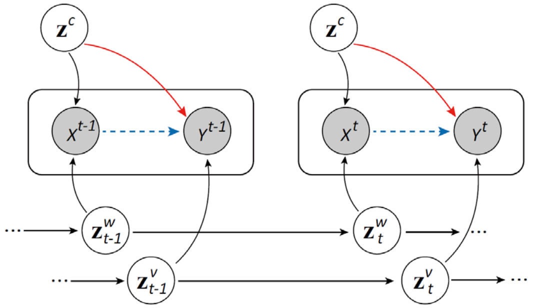
$\boldsymbol{z}^{v}_{1:T}$: dynamic domain-specific category information from Y

Methodology

- Probabilistic Generative model of LSSAE



(d)



Probabilisitic Generative Model of (d)

(1) Prior distributions:

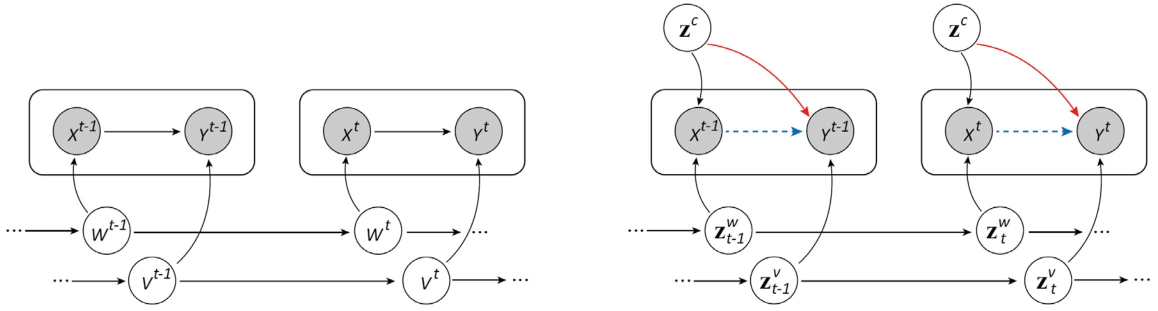
a fixed Gaussian distribution $p(\boldsymbol{z}^c) = \mathcal{N}(\mathbf{0}, \mathbf{1})$

dynamic Gaussian distribution $p(\boldsymbol{z}_t^w) = p(\boldsymbol{z}_t^w | \boldsymbol{z}_{<t}^w)$

dynamic Categorical distribution $p(\boldsymbol{z}_t^v) = p(\boldsymbol{z}_t^v | \boldsymbol{z}_{<t}^v)$

Methodology

- Probabilistic Generative model of LSSAE



(d) Probabilisitic Generative Model of (d)

(1) Prior distributions:

(2) Posterior distributions:

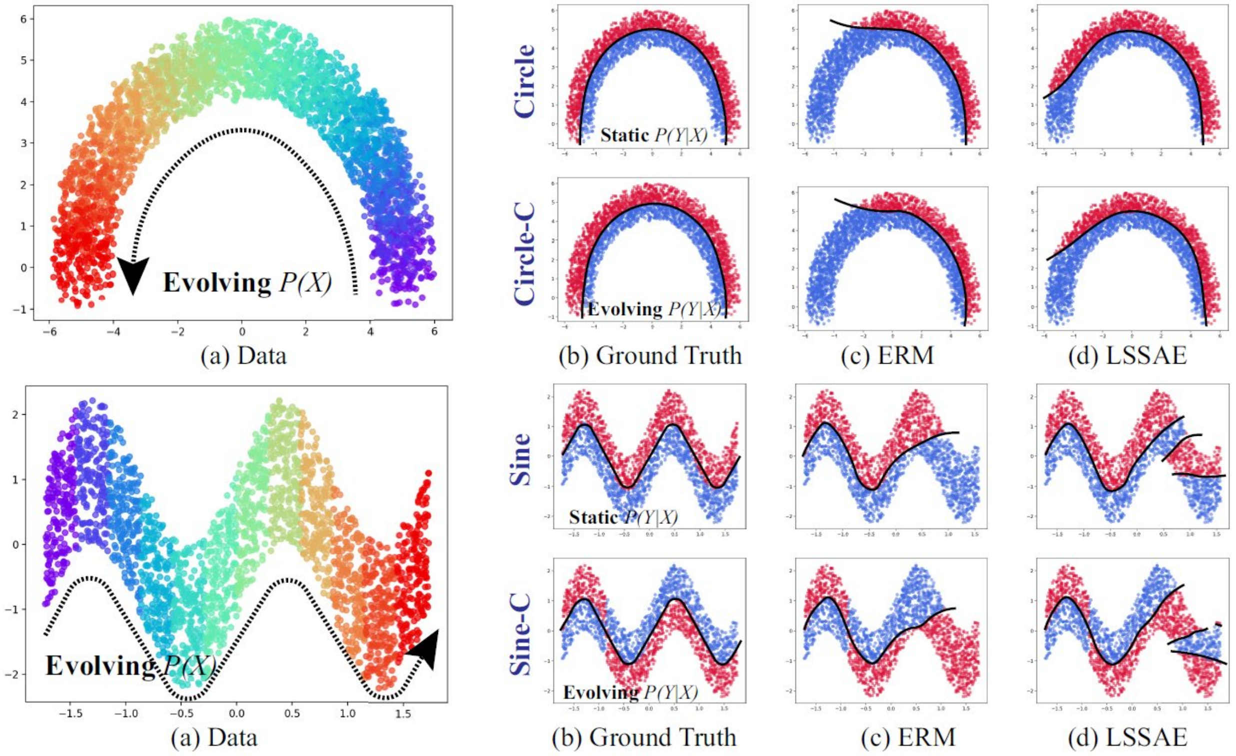
a fixed Gaussian distribution	$p(\boldsymbol{z}^c) = \mathcal{N}(\mathbf{0}, \mathbf{1})$	Align with $\mathbb{D}_{KL}$	$q(\boldsymbol{z}^c \mid \mathbf{x}_{1:T})$
dynamic Gaussian distribution	$p(\boldsymbol{z}_t^w) = p(\boldsymbol{z}_t^w \mid \boldsymbol{z}_{<t}^w)$	Align with $\mathbb{D}_{KL}$	$q(\boldsymbol{z}_t^w \mid \boldsymbol{z}_{<t}^w, \mathbf{x}_t)$
dynamic Categorical distribution	$p(\boldsymbol{z}_t^v) = p(\boldsymbol{z}_t^v \mid \boldsymbol{z}_{<t}^v)$	Align with $\mathbb{D}_{KL}$	$q(\boldsymbol{z}_t^v \mid \boldsymbol{z}_{<t}^v, \boldsymbol{y}_t)$

(3) Evidence lower bound (ELBO) for optimization:

$$\log p(\mathbf{x}_{1:T}, \boldsymbol{y}_{1:T}) \geq \sum_{t=1}^{T} \mathbb{E}_{q(\boldsymbol{z}^c, \boldsymbol{z}_t^w, \boldsymbol{z}_t^v \mid \mathbf{x}_t, \boldsymbol{y}_t)}[\log p(\mathbf{x}_t \mid \boldsymbol{z}^c, \boldsymbol{z}_t^w) p(\boldsymbol{y}_t \mid \boldsymbol{z}^c, \boldsymbol{z}_t^v)] - \lambda_1 \mathbb{D}_{KL}\big(q(\boldsymbol{z}^c \mid \mathbf{x}_{1:T}), p(\boldsymbol{z}^c)\big)$$
$$-\lambda_2 \mathbb{D}_{KL}\big(q(\boldsymbol{z}_t^w \mid \boldsymbol{z}_{<t}^w, \mathbf{x}_t), p(\boldsymbol{z}_t^w \mid \boldsymbol{z}_{<t}^w)\big) - \lambda_3 \mathbb{D}_{KL}\big(q(\boldsymbol{z}_t^v \mid \boldsymbol{z}_{<t}^v, \mathbf{x}_t), p(\boldsymbol{z}_t^v \mid \boldsymbol{z}_{<t}^v)\big)$$

Experimental Results

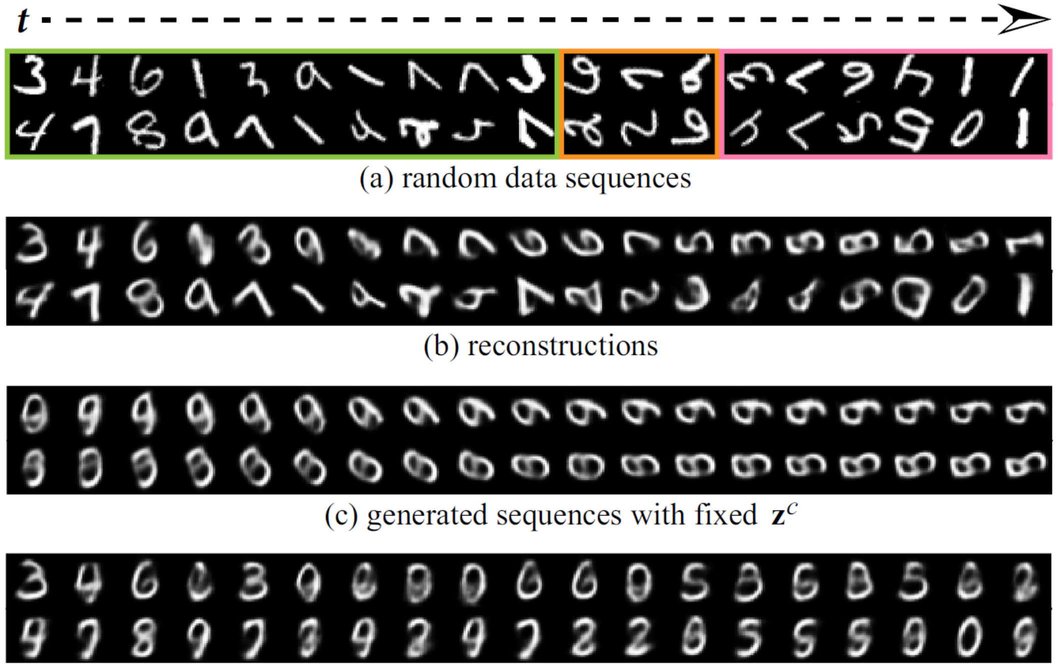
- Classification results on two toy datasets
- Generation performance on RMNIST



1. Better generalization ability than ERM
2. More suitable for *gradual* concept shift rather than *abrupt* concept shift

Experimental Results

- Classification results on two toy datasets
- Generation performance on RMNIST



(a) random data sequences

(b) reconstructions

(c) generated sequences with fixed $\mathbf{z}^c$

(d) generated sequences with fixed $\mathbf{z}_t^w$

LSSAE shows an ability of generating future unseen domains (sequence data generation, augmentation)

Thanks for Your Attention!

Contact Information:

- Tiexin Qin: tiexinqin@gmail.com
- Shiqi Wang: shiqwang@cityu.edu.hk
- Haoliang Li: haoliang.li1991@gmail.com