Variational Selective Autoencoder: Learning from Partially-Observed Heterogeneous Data

Yu Gong[1,2], Hossein Hajimirsadeghi[1], Jiawei He[1], Thibaut Durand[1], Greg Mori[1,2]

[1]Borealis AI [2]Simon Fraser University

Presenter: Yu Gong

SFU SIMON FRASER UNIVERSITY

BOREALIS AI

Motivation

- Incomplete data are often associated with heterogeneity
- Data are incomplete even during training

Background

Missingness Mechanism

$\mathbf{x_o}$ (observed data), $\mathbf{x_u}$ (unobserved data), $\mathbf{m}$ (mask)

- Missing at completely random (MCAR): $p(\mathbf{m}|\mathbf{x_o}, \mathbf{x_u}) = p(\mathbf{m}|\mathbf{x_o})$
- Missing at random (MAR): $p(\mathbf{m}|\mathbf{x_o}, \mathbf{x_u}) = p(\mathbf{m}|\mathbf{x_u})$
- Not Missing at random (NMAR): $p(\mathbf{m}|\mathbf{x_o}, \mathbf{x_u}) = p(\mathbf{m}|\mathbf{x_u})$ or $p(\mathbf{m}|\mathbf{x_o}, \mathbf{x_u})$

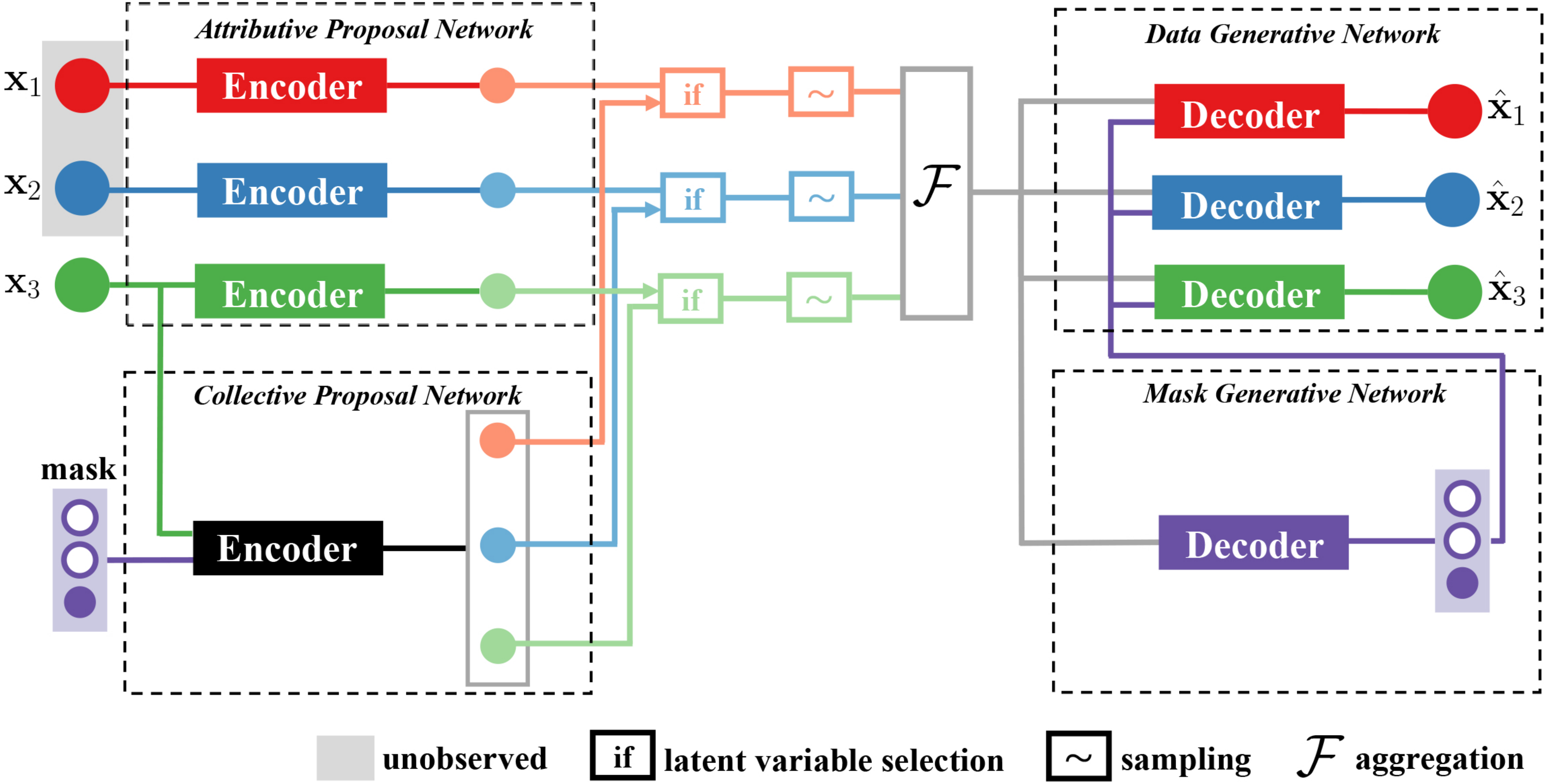
Background

Variational Autoencoder

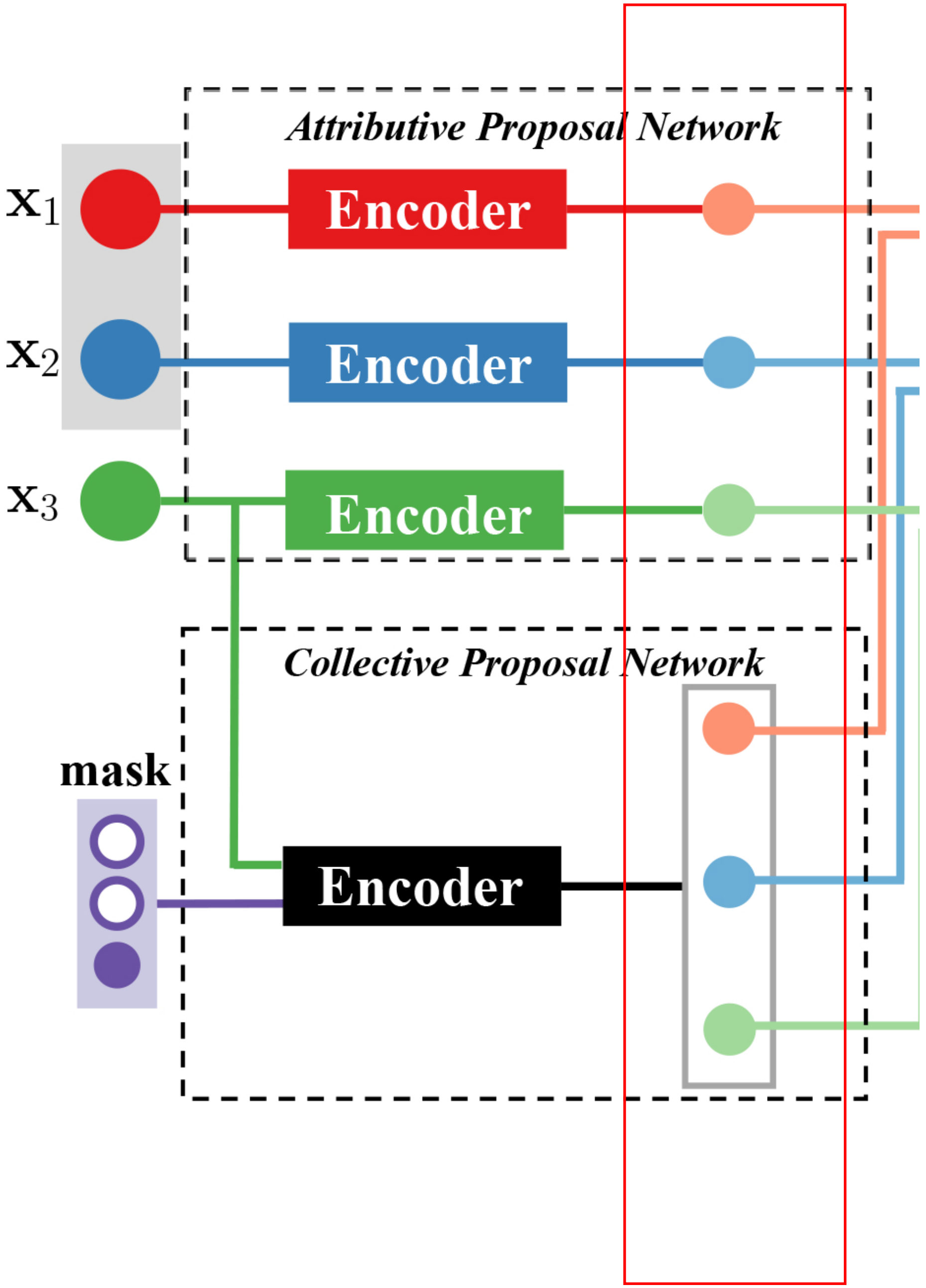
$$\log p(\mathbf{x}) \geq \mathcal{L}_{\boldsymbol{\theta},\boldsymbol{\phi}}(\mathbf{x}) = \underbrace{\mathbb{E}_{\mathbf{z}\sim q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})}[\log p_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z})]}_{\text{Conditional Log-Likelihood}} - \underbrace{D_{\mathrm{KL}}(q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})||p(\mathbf{z}))}_{\text{KL Regularizer}}$$

Proposed: Variational Selective Autoencoder

$$\log p(\mathbf{x}, \mathbf{m}) \geq \mathcal{L}_{\phi,\psi,\theta,\epsilon}(\mathbf{x}, \mathbf{m}) = \underbrace{\mathbb{E}_{\mathbf{z}\sim q_{\phi,\psi}(\mathbf{z}|\mathbf{x},\mathbf{m})}[\log p_{\theta,\epsilon}(\mathbf{x}, \mathbf{m}|\mathbf{z})]}_{\text{cond. LL}}$$
$$- \underbrace{D_{\mathrm{KL}}(q_{\phi,\psi}(\mathbf{z}|\mathbf{x}, \mathbf{m})||p(\mathbf{z}))}_{\text{KL Regularizer}},$$

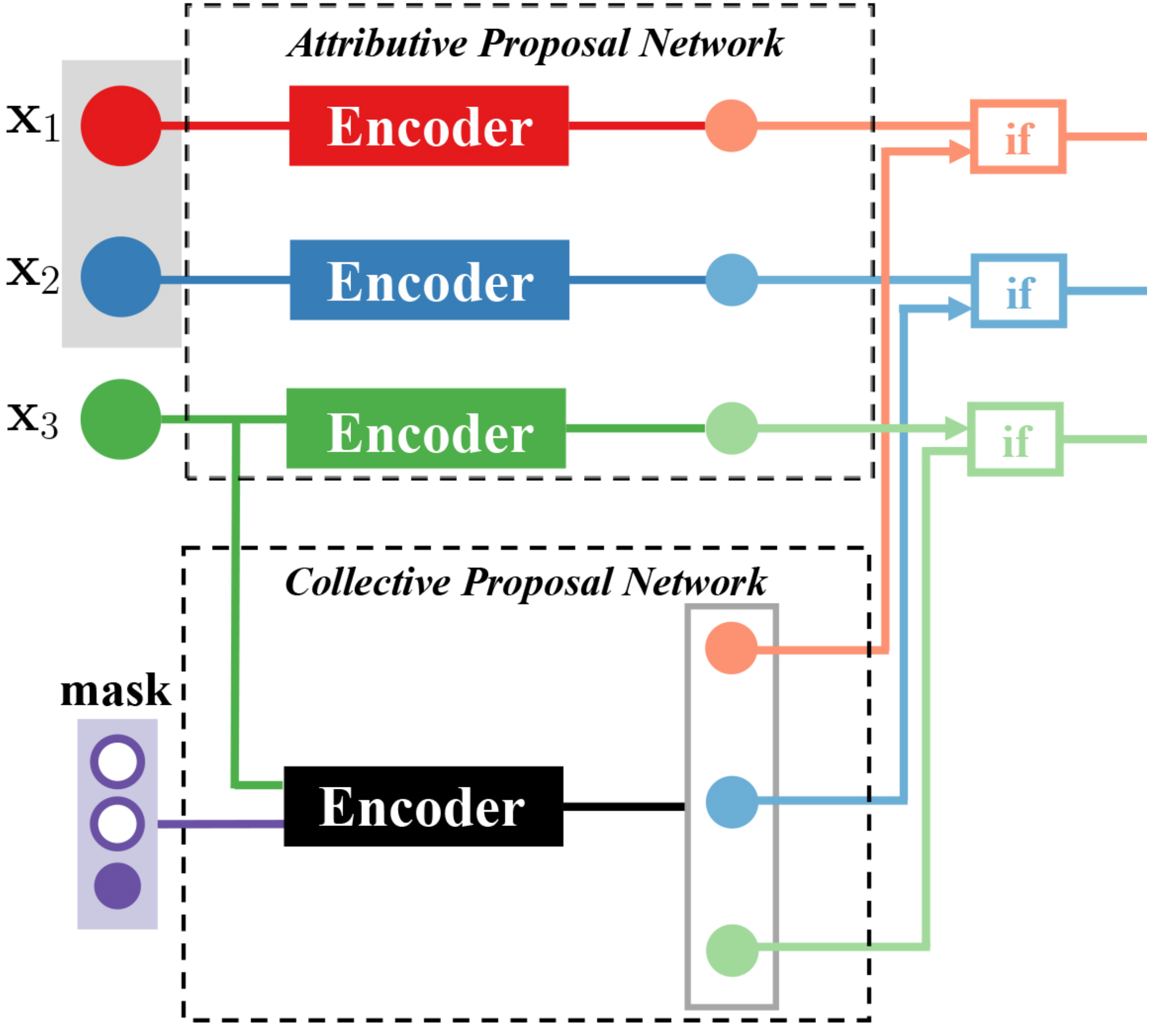


Attributive Proposal Network
Collective Proposal Network
Data Generative Network
Mask Generative Network
$\mathbf{x}_1$
$\mathbf{x}_2$
$\mathbf{x}_3$
mask
Encoder
Encoder
Encoder
Encoder
if
if
if
~
~
~
$\mathcal{F}$
Decoder
Decoder
Decoder
Decoder
$\hat{\mathbf{x}}_1$
$\hat{\mathbf{x}}_2$
$\hat{\mathbf{x}}_3$
unobserved
if latent variable selection
~ sampling
$\mathcal{F}$ aggregation



Factorization assumption

$$p(\mathbf{z}) = \prod_{i=1}^{M} p(\mathbf{z}_i), \qquad q(\mathbf{z}|\mathbf{x}, \mathbf{m}) = \prod_{i=1}^{M} q(\mathbf{z}_i|\mathbf{x}, \mathbf{m}).$$



Selective Proposal Distribution

$$q_{\phi,\psi}(\mathbf{z}_i|\mathbf{x},\mathbf{m}) = \begin{cases} q_\phi(\mathbf{z}_i|\mathbf{x}_i) & \text{if } m_i = 1 \\ q_\psi(\mathbf{z}_i|\mathbf{x_o},\mathbf{m}) & \text{if } m_i = 0 \end{cases}$$

Expected ELBO over unobserved attributes

$$\mathcal{L}'_{\boldsymbol{\phi},\boldsymbol{\psi},\boldsymbol{\theta},\boldsymbol{\epsilon}}(\mathbf{x_o},\mathbf{m}) = \mathbb{E}_{\mathbf{x_u}}[\mathcal{L}_{\boldsymbol{\phi},\boldsymbol{\psi},\boldsymbol{\theta},\boldsymbol{\epsilon}}(\mathbf{x_o},\mathbf{x_u},\mathbf{m})]$$

Experiment

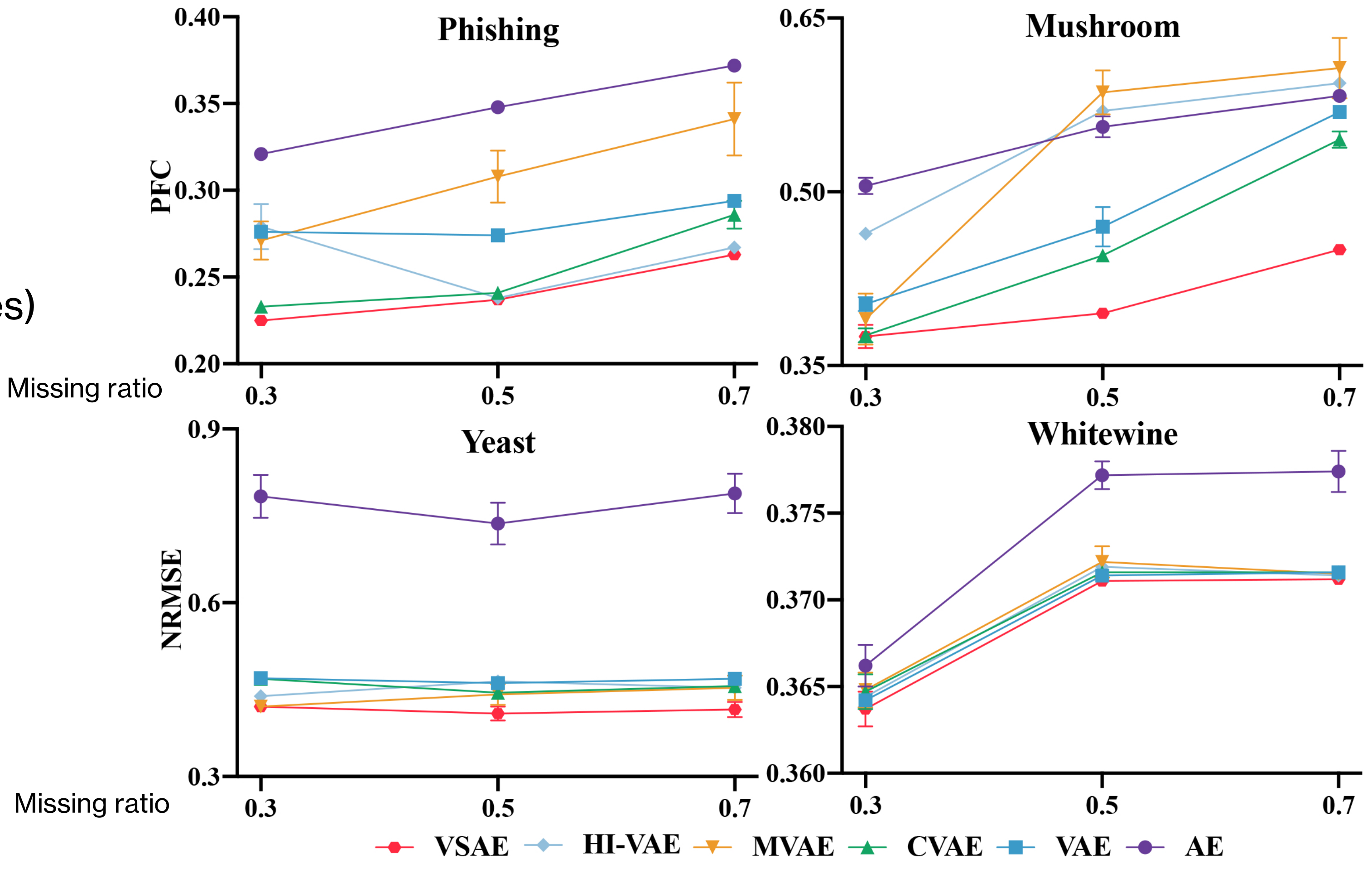
Data imputation under MCAR setting

Dataset: *UCI repository*

Metric:

PFC (portion of falsely classified attributes)

NRMSE (normalized RMSE)



Experiment

Data imputation under non-MCAR setting

Dataset: *Yeast, Whitewine*

Metric: *NRMSE* (normalized RMSE)

	Method	MAR	NMAR
Yeast	MIWAE	$\mathbf{0.475 \pm 0.005}$	0.456 ± 0.036
	VSAE (ours)	$\mathbf{0.472 \pm 0.006}$	$\mathbf{0.425 \pm 0.007}$
Whitewine	MIWAE	$0.3834 \pm \Delta$	$0.3723 \pm \Delta$
	VSAE (ours)	$\mathbf{0.3825 \pm \Delta}$	$\mathbf{0.3717 \pm \Delta}$

Synthesize non-MCAR:
***MAR*:** 25% attributes are default observed, then sample the remaining mask from sigmoid function.
NMAR: sample the mask for each attribute by applying sigmoid.

Experiment

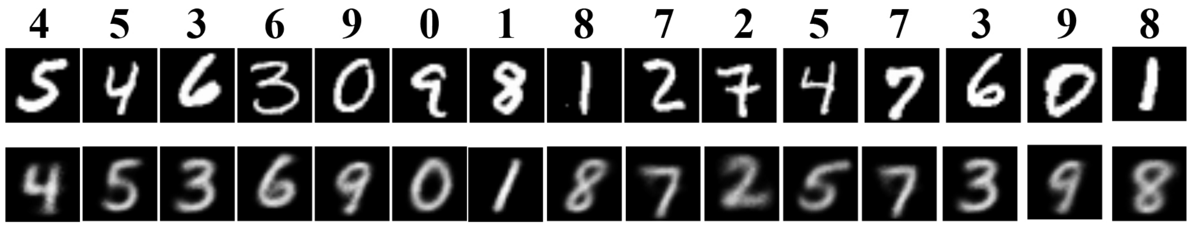
Dataset: *MNIST + MNIST bimodal*

Metric: *MSE*

Synthesize multi-modal dataset:
MNIST as {(0,9),(1,8),(2,7),(3,6),(4,5)}
.

Data generation

Data imputation given one attribute



Experiment

Dataset: *Fashion MNIST, CMU-MOSI*

Metric: *MSE*

	FashionMNIST + label (PFC)		**MNIST + MNIST**		**CMU-MOSI**		
Attribute type	Image	Label	Digit–1	Digit–2	Text	Audio	Image
AE	0.1105 ± 0.001	0.366 ± 0.01	$0.1077 \pm \Delta$	$0.1070 \pm \Delta$	0.035 ± 0.003	0.224 ± 0.025	0.019 ± 0.003
VAE	$0.0885 \pm \Delta$	0.411 ± 0.01	$0.0734 \pm \Delta$	$0.0682 \pm \Delta$	$0.034 \pm \Delta$	0.202 ± 0.003	$\mathbf{0.017 \pm \Delta}$
CVAE w/ mask	$0.0887 \pm \Delta$	0.412 ± 0.01	$0.0733 \pm \Delta$	$0.0679 \pm \Delta$	$0.043 \pm \Delta$	0.257 ± 0.002	$0.020 \pm \Delta$
MVAE	0.1402 ± 0.026	0.374 ± 0.07	$0.0760 \pm \Delta$	$0.0802 \pm \Delta$	$0.44 \pm \Delta$	0.213 ± 0.001	$0.025 \pm \Delta$
HI-VAE	0.1575 ± 0.006	0.405 ± 0.01	$0.0772 \pm \Delta$	$0.0725 \pm \Delta$	$0.047 \pm \Delta$	0.211 ± 0.005	$0.0267 \pm \Delta$
VSAE (ours)	$\mathbf{0.0874 \pm \Delta}$	$\mathbf{0.356 \pm 0.01}$	$\mathbf{0.0712 \pm \Delta}$	$\mathbf{0.0663 \pm \Delta}$	$\mathbf{0.033 \pm \Delta}$	$\mathbf{0.200 \pm \Delta}$	$\mathbf{0.017 \pm \Delta}$