Causal generative models and representation learning[1]

Xinwei Shen

Seminar for Statistics and AI Center, ETH Zurich

ETH-UCPH-TUM Workshop

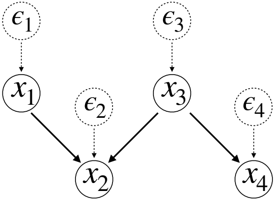
October 11, 2022

[1] Shen, Xinwei, et al. "Weakly Supervised Disentangled Generative Causal Representation Learning." *Journal of Machine Learning Research* 23 (2022): 1-55.

Causality and machine learning

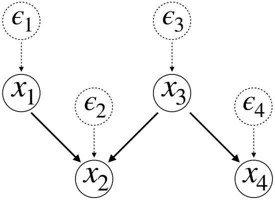
Traditional causality

ML

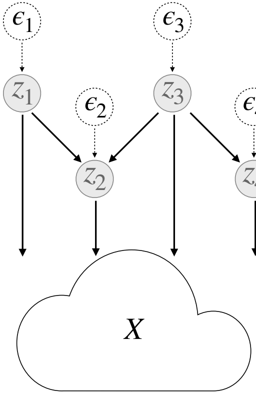


Causality and machine learning

Traditional causality

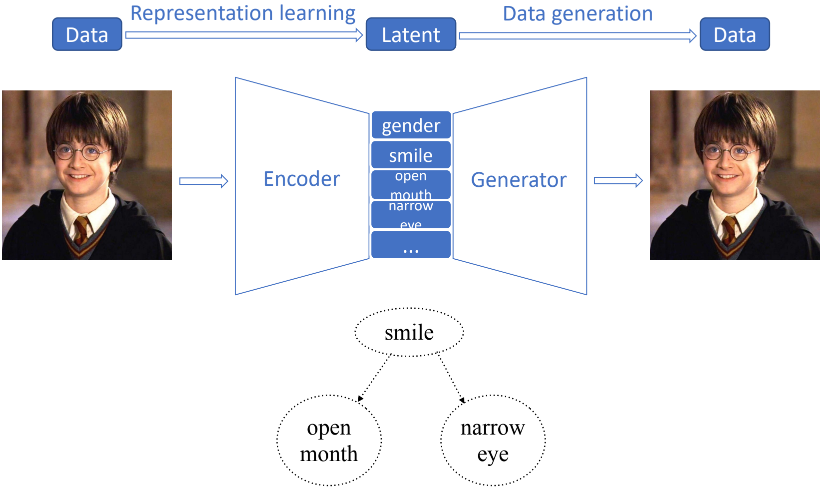


ML



e.g., images

Representation learning and generative models



Methods for representation learning and generative models

Notations

- Observed data $x \sim p_*$ on $\mathcal{X} \subseteq \mathbb{R}^d$
- Latent variable $z \sim p_z$ on $\mathcal{Z} \subseteq \mathbb{R}^k$
- Goal: to learn an *encoder* $E_\phi : \mathcal{X} \to \mathcal{Z}$ and a *generator* $G_\theta : \mathcal{Z} \to \mathcal{X}$.

Methods for representation learning and generative models

Notations

- Observed data $x \sim p_*$ on $\mathcal{X} \subseteq \mathbb{R}^d$
- Latent variable $z \sim p_z$ on $\mathcal{Z} \subseteq \mathbb{R}^k$
- Goal: to learn an *encoder* $E_\phi : \mathcal{X} \to \mathcal{Z}$ and a *generator* $G_\theta : \mathcal{Z} \to \mathcal{X}$.

Methods

- VAE

$$\max_{\phi,\theta} \mathbb{E}_{x \sim p_*}[\ln q_\phi(z|x) - D_{\mathrm{KL}}(q_\phi(z|x), p_z(z))] \tag{1}$$

- GAN

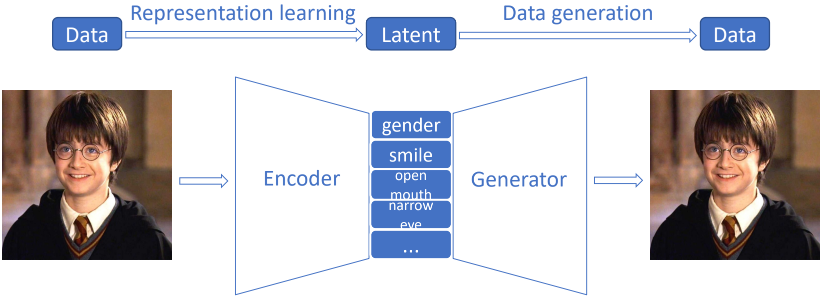
$$\min_{\phi,\theta} \max_{D} [\mathbb{E}_{x \sim p_*, z \sim q_\phi(z|x)}(\ln D(x,z)) + \mathbb{E}_{z \sim p_z, x \sim p_\theta(x|z)}(1 - \ln D(x,z))] \tag{2}$$

Disentanglement

- *Disentanglement*: each dimension of the latent variable measures a distinct generative factor of the data (Bengio et al., 2013).

Disentanglement

- *Disentanglement*: each dimension of the latent variable measures a distinct generative factor of the data (Bengio et al., 2013).



Previous work

- β-VAE (Higgins et al., 2017)

$$\max_{\phi,\theta} \mathbb{E}_{x\sim p_*}[\ln q_\phi(z|x) - \beta D_{\mathrm{KL}}(q_\phi(z|x), p_z(z))] \tag{3}$$

with $\beta > 1$.

Previous work

- β-VAE (Higgins et al., 2017)

$$\max_{\phi,\theta} \mathbb{E}_{x \sim p_*}[\ln q_\phi(z|x) - \beta D_{\mathrm{KL}}(q_\phi(z|x), p_z(z))] \quad (3)$$

  with $\beta > 1$.
- Problems:
  - Independence assumption: the underlying factors are mutually independent.

Previous work

- β-VAE (Higgins et al., 2017)

$$\max_{\phi,\theta} \mathbb{E}_{x\sim p_*}[\ln q_\phi(z|x) - \beta D_{\mathrm{KL}}(q_\phi(z|x), p_z(z))] \tag{3}$$

  with $\beta > 1$.
- Problems:
  - Independence assumption: the underlying factors are mutually independent.
    $\rightarrow$ what if the true factors are causally related?

Previous work

- β-VAE (Higgins et al., 2017)

$$\max_{\phi,\theta} \mathbb{E}_{x\sim p_*}[\ln q_\phi(z|x) - \beta D_{\mathrm{KL}}(q_\phi(z|x), p_z(z))] \tag{3}$$

  with $\beta > 1$.
- Problems:
  - Independence assumption: the underlying factors are mutually independent.
    $\rightarrow$ what if the true factors are causally related?
  - Unidentifiability of true latent variables

Previous work

- β-VAE (Higgins et al., 2017)

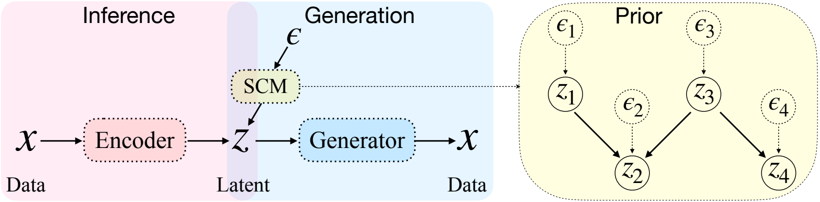
$$\max_{\phi,\theta} \mathbb{E}_{x\sim p_*}[\ln q_\phi(z|x) - \beta D_{\mathrm{KL}}(q_\phi(z|x), p_z(z))] \tag{3}$$

  with $\beta > 1$.
- Problems:
  - Independence assumption: the underlying factors are mutually independent.
    $\rightarrow$ what if the true factors are causally related?
  - Unidentifiability of true latent variables
    $\rightarrow$ Locatello et al. (2019) showed that unsupervised learning of disentangled representations is impossible.

Causal disentanglement learning

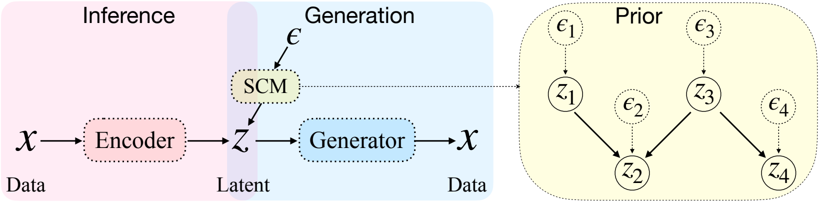
Model

- Using a structural causal model (SCM) as the prior distribution of z.



Model

- Using a structural causal model (SCM) as the prior distribution of z.



- Prior distribution $p_\beta(z)$, where parameter β includes the causal structure and structural equations.

Formulation and algorithm

- Formulation with weak supervision

$$\min_{\theta,\phi,\beta} \left[D_{\mathrm{KL}}(q_\phi(x,z), p_{\theta,\beta}(x,z)) + \lambda \mathbb{E}[c(E(X), Y)] \right]$$

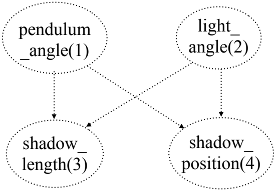
- We adopt our proposed efficient GAN algorithm for optimization (Shen et al., 2022).
- Identifiability and statistical consistency.

Experimental results

Data sets

- Synthesized dataset Pendulum (Yang et al., 2020)

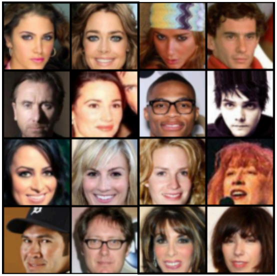
(a)



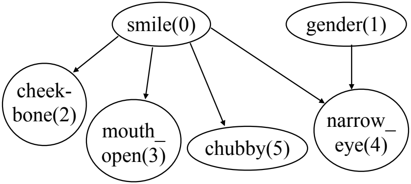
(b)

Data sets

- CelebA (Liu et al., 2015)
- Meta-data: some labeled binary attributes.



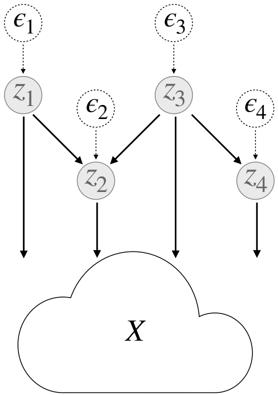
(a)



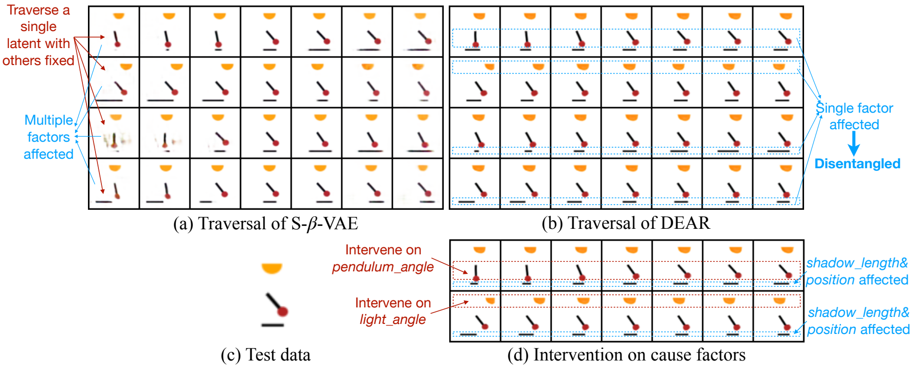
(b)

Interventional generation

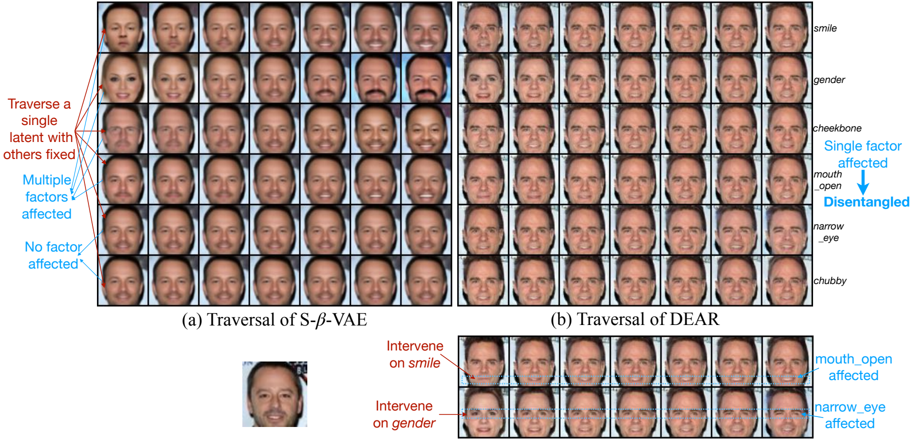
- Standard traversals: direct causal effect
- Do-interventions on causes, $\mathbb{P}_{\beta}^{\mathrm{do}(z_i=c)}(z)$: total causal effects



Interventional generation



Interventional generation



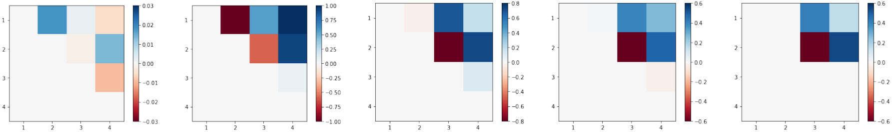
(a) Traversal of S-β-VAE

(b) Traversal of DEAR

(c) Test data

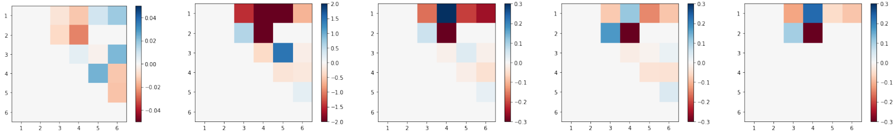
(d) Intervention on cause factors

Structure learning (Pendulum)



(a) Epoch 0 (b) Epoch 100 (c) Epoch 200 (d) Epoch 500 (e) True

Structure learning (CelebA)



(f) Epoch 0 (g) Epoch 5 (h) Epoch 50 (i) Epoch 150 (j) True

Discussion on prediction based on causal disentanglement

Another important application is “better” prediction based on the causal disentangled representations.

[2] Nonlinear prediction

Discussion on prediction based on causal disentanglement

Another important application is “better” prediction based on the causal disentangled representations.

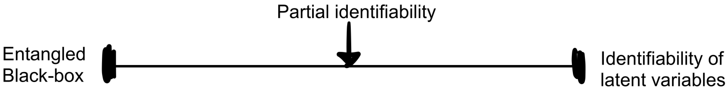
- Disentanglement $\Rightarrow$ interpretability

[2] Nonlinear prediction

Discussion on prediction based on causal disentanglement

Another important application is “better” prediction based on the causal disentangled representations.

- Disentanglement $\Rightarrow$ interpretability

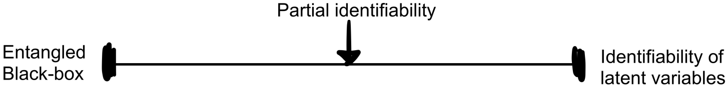


[2]Nonlinear prediction

Discussion on prediction based on causal disentanglement

Another important application is “better” prediction based on the causal disentangled representations.

- Disentanglement $\Rightarrow$ interpretability



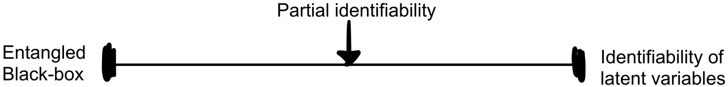
- Causality $\Rightarrow$ robustness[2]

[2] Nonlinear prediction

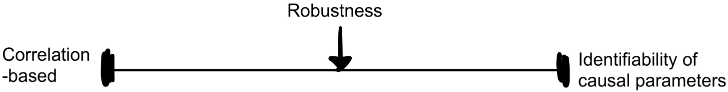
Discussion on prediction based on causal disentanglement

Another important application is “better” prediction based on the causal disentangled representations.

- Disentanglement $\Rightarrow$ interpretability

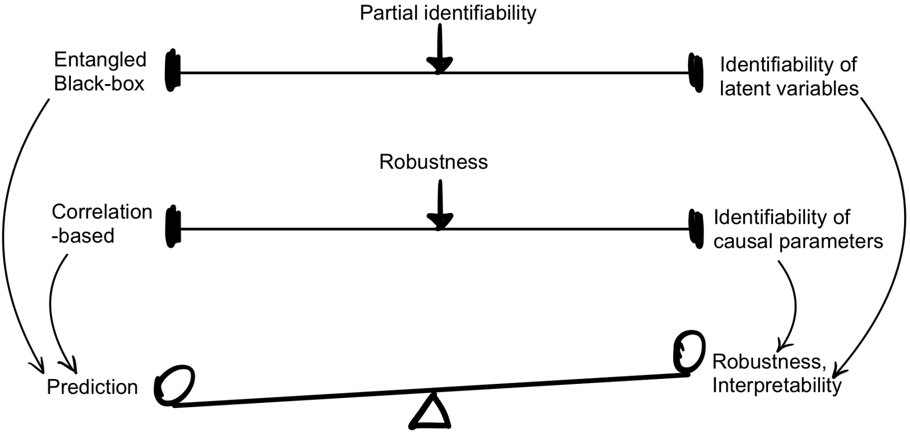


- Causality $\Rightarrow$ robustness[2]

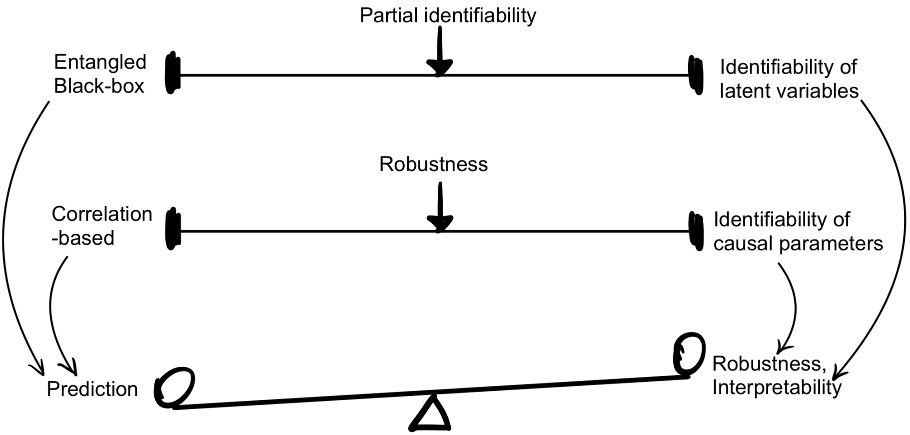


[2] Nonlinear prediction

Discussion on prediction based on causal disentanglement



Discussion on prediction based on causal disentanglement



- Ambition: interpretability and robustness in a joint formulation

Summary

- Causality and machine learning
  - causal generative models and representation learning
- Causal disentanglement learning
  - an SCM as the prior distribution for the latent variable
  - interventional generation
- Trade-off between prediction and interpretability and robustness

References

Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, *35*(8), 1798–1828.

Higgins, I., Matthey, L., Pal, A., Burgess, C., Glorot, X., Botvinick, M., ... Lerchner, A. (2017). beta-vae: Learning basic visual concepts with a constrained variational framework. In *Iclr*.

Liu, Z., Luo, P., Wang, X., & Tang, X. (2015). Deep learning face attributes in the wild. In *Proceedings of the ieee international conference on computer vision* (pp. 3730–3738).

Locatello, F., Bauer, S., Lucic, M., Raetsch, G., Gelly, S., Schölkopf, B., & Bachem, O. (2019, June). Challenging common assumptions in the unsupervised learning of disentangled representations. In *Proceedings of the 36th international conference on machine learning (icml)* (Vol. 97, pp. 4114–4124). PMLR. Retrieved from `http://proceedings.mlr.press/v97/locatello19a.html`

Shen, X., Chen, K., & Zhang, T. (2022). Asymptotic statistical analysis of f-divergence gan. *arXiv preprint arXiv:2209.06853*.

Yang, M., Liu, F., Chen, Z., Shen, X., Hao, J., & Wang, J. (2020). Causalvae: Structured causal disentanglement in variational autoencoder. *arXiv preprint arXiv:2004.08697*.

Thanks