Latent space of VAE for DAG learning

김당찬, 강병국, 김재석, 김민찬

November 9, 2023

1 Research Objective

본 연구의 목적은 VAE(Variational AutoEncoder)를 기반으로 한 GNN(Graph neural network)의 잠재공간(latent space)에 대한 통계적 접근 및 해석 방법을 찾는 것이다. 일반적으로 linear SEM 가정에서 VAE를 이용하면 노이즈 행렬을 잠재변수(latent variable)로 간주할 수 있다. 이때 노이즈 확률변수가 잠재변수에 대응되기 때문에, Gaussian 노이즈의 경우 정규사전분포를 사용할 수 있으나 non-Gaussian 분포를 따르는 경우 사전분포 설정이 어렵다는 문제가 존재한다.

이에, 본 연구에서는 잠재공간의 구조와 분포에 대한 통계적 해석을 진행한다. 구체적으로 VAE에서의 reparametrization trick을 이용하기 위해 non-Gaussian prior에 대한 surrogate distribution 혹은을 이용한다. 이를 통해 적절한 approximated prior distribution를 도출하여 non-Gaussian linear SEM을 학습하는 방법을 제안하고자 한다. 또한, 이를 바탕으로 랜덤 생성된 non-Gaussian DAG에 대한 시뮬레이션과 document 데이터 등의 실제 데이터에 대한 실험을 진행한다.

2 Preliminary Results

2.1 DAG and Linear SEM

DAG $\mathcal{G} = (V, E)$가 $|V| = m$ 개의 노드를 갖고 가중인접행렬(weighted adjacency matrix) $A \in \mathbb{R}^{m\times m}$을 갖는다고 가정하자. m개의 각 노드는 random variable에 대응되는데, 이때 faithful joint distribution으로부터 얻은 표본을 $X \in \mathbb{R}^{m\times d}$ 라고 하자. Linear SEM에서는 그래프 구조에 대해 다음을 가정한다.

$$X = A^T X + Z \tag{1}$$
$$= (I - A^T)^{-1} Z. \tag{2}$$

행렬 $Z \in \mathbb{R}^{m\times d}$는 노이즈 행렬을 나타낸다. DAG에서는 topological sorting으로 인접행렬이 상삼각행렬이 되므로, 두번째 식이 성립한다.

DAG 학습은 DAG의 구조를 주어진 데이터로부터 재구성하는 것이다. 그러나 이는 NP-hard problem에 해당한다. [1] 이로 인해 다양한 가정을 기반으로 한 DAG 학습 알고리즘들이 존재한다. 학습 알고리즘에는 크게 Constraint-based methods [2] 와 Score-based methods [3] 가 있다. 또한 보통 이런 방법들은 변수 사이의 functional form을 가정하고 학습하는데, 위의 경우처럼 (식 2) linear SEM을 가정할 수 있다 [4]. Score-based methods는 DAG의 search space가 너무 큰 combinatorial problem, 즉 노드의 개수 증가에 따라 score를 고려해야 하는 DAG의 갯수가 superexponential하게 증가하는 문제가 있었는데, 근래 이를 continuous optimization 문제로 해결한 방법도 제안되었다 [5]. 이를 토대로 DAG learning을 Deep learning Framework에서 진행할 수 있게 되었다.

2.2 Variational Autoencoder

VAE[6]에서는 데이터 x가 잠재변수 z로부터 생성된다고 가정한다. 입력변수 x의 분포 $p_\theta(x)$나, 잠재변수 z의 사후분포 $p_\theta(z|x)$는 직접 계산하기 어렵기에(intractable), true posterior $p_\theta(z|x)$를 근사하는 다른 분포족

$q_\phi(z|x)$를 도입한다. 이로부터 x의 가능도함수는 다음과 같이 표현된다.

$$\log p_\theta(x) = D_{KL}(q_\phi(z|x)||p_\theta(z|x)) - D_{KL}(q_\phi(z|x)||p_\theta(z)) + \mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x|z)] \tag{3}$$

이로부터 VAE의 목적함수인 ELBO(evidence lower bound)는 다음과 같이 정의된다. VAE에서는 ELBO를 최대화하여 로그가능도를 최대화하고, 사후분포와 그 근사분포의 KL 발산을 최소화한다.

$$\mathcal{L}(\theta, \phi, x) := -D_{KL}(q_\phi(z|x)||p_\theta(z)) + \mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x|z)]$$

reparametrization trick VAE와 같은 stochastic variational inference에서는 reparametrization trick으로 [7] ϕ에 대한 gradient 계산을 간접적으로 수행한다. 조건부분포 $q_\phi(z|x)$ 로부터 표본을 추출하는 대신, 확률변수 $\epsilon \sim p(\epsilon)$을 도입해 다음과 같이 z를 deterministic, differentiable $z = g_\phi(\epsilon, x)$ 로 reparametrization 하여 나타낸다.

$$\mathbb{E}_{q_\phi(z|x)}[-\log q_\phi(z|x) + \log p_\theta(x, z)] = \mathbb{E}_{p(\epsilon)}[-\log q_\phi(g_\phi(\epsilon, x)|x) + \log p_\theta(x, g_\phi(\epsilon, x))] \tag{4}$$

Latent space of VAE VAE에서 정의되는 잠재공간는 데이터의 representation을 저차원의 잠재변수로 표현한 것이다. 인코더에 대한 특정 approximation과 디코더에 대한 orthogonality 제약조건을 하에서 PCA와 VAE가 유사한 latent spcae를 갖는다는 것이 알려져 있다. [8]

2.3 Graph Neural Network

Linear SEM 식 (2)를 $X = f_A(Z)$ 라고 표기하면, 이는 일반적인 GNN의 기본 구조가 되며 나아가 다음과 같은 generalized linear SEM도 고려할 수 있다. [9]

$$X = f_2((I - A^T)^{-1} f_1(Z)) \tag{5}$$

f_1, f_2에 신경망 구조를 가정하면, 이는 autoencoder 형태의 모형이 된다.

Latent variable for Graph data

Graph structured data를 VAE와 같이 잠재공간 기반의 모델로 다루기 위해서는 잠재변수에 대한 설정이 필요하다. Table 1은 graph structured data를 다룬 여러 VAE 선행연구들에서 잠재공간의 구조가 어떤 형태로 가정되었는지 정리한 표이다.

Model	**Latent variable**	**Prior**
DAG-GNN[9]	Noise matrix from Linear SEM, $Z \in \mathbb{R}^{m\times d}$	$\mathcal{MN}_{m,d}(0, I, I)$
VGAE[10]	stochastic latent variable, $Z \in \mathbb{R}^{n\times d}$	$\mathcal{N}(0, I)$
CausalVAE[11]	Endogeneous latent variable $\mathbf{z} \in \mathbb{R}^{\mathbf{n}}$ Exogeneous latent variable $\mathbf{u} \in \mathbb{R}^{\mathbf{n}}$	$\mathbf{z} = \mathbf{A}^{\mathbf{T}}\mathbf{z} + \mathbf{u}$ $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
Dirichlet Graph VAE[12]	cluster membership of each node, Z_i	$Z_i \sim \mathcal{D}(\alpha)$

Table 1: Different latent structure of VAEs for graph-structured data

3 Experiment

실험 과정은 simulation study와 application study를 진행한다. Simulation에서는, Erdös-Rényi model [13]로 random DAG에 대한 데이터를 생성하여, 제안된 모델을 바탕으로 학습 및 평가를 진행한다. 문서 데이터에는 GNN 기반 학습 방법이 다수 제안된 바 있다. [14, 15, 16, 17] 다만, citation network [18] 와 같이 text(document) 사이의 관계를 나타내는 그래프 데이터의 경우, DAG 기반의 분석이 가능하다. [19] 이러한 데이터에 대해 제안된 모델을 적용해 보고자 한다. 모델의 성능 평가의 경우 SHD(Structural Hamming Distance)와 FDR(False Discovery Rate)를 사용한다. [9]

References

[1] D. M. Chickering, C. Meek, and D. Heckerman, "Large-sample learning of bayesian networks is np-hard," *CoRR*, vol. abs/1212.2468, 2012.

[2] P. Spirtes, C. Glymour, and R. Scheines, *Causation, Prediction, and Search.* MIT press, 2nd ed., 2000.

[3] D. M. Chickering, "Optimal structure identification with greedy search," *J. Mach. Learn. Res.*, vol. 3, pp. 507–554, mar 2003.

[4] A. Ghoshal and J. Honorio, "Learning linear structural equation models in polynomial time and sample complexity," *CoRR*, vol. abs/1707.04673, 2017.

[5] X. Zheng, B. Aragam, P. Ravikumar, and E. P. Xing, "Dags with no tears: Continuous optimization for structure learning," 2018.

[6] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[7] M. Figurnov, S. Mohamed, and A. Mnih, "Implicit reparameterization gradients," 2019.

[8] M. Rolinek, D. Zietlow, and G. Martius, "Variational autoencoders pursue pca directions (by accident)," 2019.

[9] Y. Yu, J. Chen, T. Gao, and M. Yu, "Dag-gnn: Dag structure learning with graph neural networks," 2019.

[10] T. N. Kipf and M. Welling, "Variational graph auto-encoders," 2016.

[11] M. Yang, F. Liu, Z. Chen, X. Shen, J. Hao, and J. Wang, "Causalvae: Disentangled representation learning via neural structural causal models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9593–9602, June 2021.

[12] J. Li, T. Yu, J. Li, H. Zhang, K. Zhao, Y. Rong, H. Cheng, and J. Huang, "Dirichlet graph variational autoencoder," 2020.

[13] P. ERDdS and A. R&wi, "On random graphs i," *Publ. math. debrecen*, vol. 6, no. 290-297, p. 18, 1959.

[14] N. Peng, H. Poon, C. Quirk, K. Toutanova, and W. tau Yih, "Cross-sentence n-ary relation extraction with graph lstms," 2017.

[15] D. Marcheggiani and I. Titov, "Encoding sentences with graph convolutional networks for semantic role labeling," 2017.

[16] M. Schlichtkrull, T. N. Kipf, P. Bloem, R. van den Berg, I. Titov, and M. Welling, "Modeling relational data with graph convolutional networks," 2017.

[17] Y. Zhang, X. Yu, Z. Cui, S. Wu, Z. Wen, and L. Wang, "Every document owns its structure: Inductive text classification via graph neural networks," 2020.

[18] J. R. Clough and T. S. Evans, "What is the dimension of citation space?," *Physica A: Statistical Mechanics and its Applications*, vol. 448, pp. 235–247, 2016.

[19] J. Wu, Z. Xuan, and D. Pan, "Enhancing text representation for classification tasks with semantic graph structures," *International Journal of Innovative Computing, Information and Control*, vol. 7, 05 2011.