Semi-Implicit Graph Variational Auto-Encoder

NeurIPS 2019에 실렸다

1. Abstract

Semi-implicit graph variational auto-encoder (SIG-VAE)는 VGAE가 graph data를 모델링할 수 있는 측면에서의 flexibility를 늘리기 위해 제안된 구조이다. SIG-VAE는 hierarchical variational framework과 Bernoulli-Poisson link decoder를 도입해서 neighboring node sharing을 가능하게 하고, graph dependency structure의 더 나은 generative modeling을 수행한다. SIG-VAE는 heavy tail, multiple modes, skewness, rich dependency structure와 같은 feature를 가지고 있는 graph의 implicit posterior를 충실하게 모델링할 수 있는 semi-implicit한 hierarchical variational inference가 가능할 뿐만 아니라, 이 hierarchical construction은 실존하는 graph property를 더 잘 알아내기 위해서 더 flexible한 generative graph model을 우리에게 제공한다. SIG-VAE는 실제 sparse graph를 모델링하는 데 적합하도록 조심스럽게 디자인한 generative model과, complex posterior를 알아낼 수 있도록 graph structural information과 distribution uncertainty에서 퍼져나가는 정교한 variational inference network를 통합한다. SIG-VAE는 분명하게 단순히 VGAE+Semi-Implicit Variational Inference (SIVI)나 VGAE + Normalizing Flow (NF)를 능가하는데, 이것은 이 모델들이 inference network 안에서 uncertainty를 propagate하지 않고, 우리의 모델이 VGAE보다 더 interpretable한 latent representation을 제공하기 때문이다. 다양한 graph data를 이용한 실험은 SIG-VAE가 SOTA method를 여러 graph analytic task에서 능가했음을 보여준다.

1. Introduction

Graph data를 분석하는 것은 다양하게 활용될 수 있는 중요한 ML task이다. Transportation networks, social networks, gene co-expression networks, recommendation systems가 graph로 모델링될 수 있는 몇 가지 예시이다. 여기서 각 node는 agent (ex. 교차로, 사람, 유전자)에 해당하고 edge는 agent 사이의 interaction을 관장한다. Link prediction, clustering, node classification 등으로 graph dataset을 분석하려는 challenge는 graph structural information을 모델에서 어떻게 적용할지가 될 것이다. Graph representation learning은 graph structural information을 low-dimensional latent space 안에 있는 feature vector로 나타내고, 이걸 downstream analytic task에 다시 이용하는 것을 목표로 한다.

현존하는 여러 방식이 각 node가 latent space 안의 deterministic한 한 점에 embed된다고 가정하는 반면에, uncertainty를 모델링하는 것은 여러 분야에서 중요한 의미를 가진다. 예를 들어서, Knowledge Graph에서 link prediction이 비싼 약학적 실험을 진행하는 데 쓰인다면, prediction에서 모델의 confidence level이 어떻게 되는지를 안다면 유익할 것이다. 이 문제를 해결하기 위해서, VGAE는 각 node를 latent space의 random variable에 embed한다. VGAE의 많은 인기에도 불구하고, 1) Variational distribution에 적용되는 Gaussian assumption은 graph가 주어질 시 true posterior distribution이 분명하게 Gaussian assumption을 위반할 시 variational inference의 flexibility가 제한된다. 2) Inner-product decoder가 generative model flexibility를 제한한다. 최근의 연구들이 prior distribution을 바꿔봄으로써 첫 번째 문제를 해결하려고 노력하지만, 실전적인 성공을 거뒀다고는 보기 힘들다. 심지어 후자는 우리의 지식 수준으로는 잘 연구되지 않았다.

최근에 개발된, flexible posterior distribution과 effective한 optimization 사이의 흥미로운 combination을 제시한 Semi-Implicit Variational Inference (SIVI)와 Normalizing Flow (NF)에 영감을 얻어서, 우리는 graph structured data의 node embedding을 위한 hierarchical variational graph framework을 제시하고, latent space의 각 node에서 posterior distribution의 expressiveness를 눈에 띄게 늘렸다. SIVI는 mean-field variational inference[[1]](#footnote-1)를, flexible (implicit) mixing distribution으로 강화했다. NF는 simple Gaussian random variable을, tractable Jacobian을 가지는 일련의 invertible한 미분가능한 함수들을 통해서 transform시킨다. NF는 hierarchy 안의 mixing distribution으로 하여금 explicit한 probability density function을 가지도록 제한하지만, SIVI는 그런 제한을 두지는 않는다. SIVI와 NF 모두 복잡한 posterior distribution을 모델링할 수 있고, 이것은 underlying true embedded node distribution이 heavy tail이나 multiple mode와 같은 성질을 보일 때 유용할 것이다. 우리는 이따가 graph neighborhood information을 integrate하지 않는다면, 단순히 SIVI/NF와 VGAE를 결합하는 것만으로는 posterior distribution을 완전히 exploit할 수 없다는 것을 주장할 것이다. 반면에, 두 번째 VGAE 문제점으로 제시된 generative model의 flexibility를 해결하지 못하기도 한다.

방금 언급한 문제를 해결하기 위해서, 이전 연구에서처럼 posterior distributon family를 explicit하게 선택하는 대신에, 우리의 hierarchical variational framework은 단순한 형태의 optimization 문제 형태를 유지하면서 implicit한 posterior를 학습할 수 있는 stochastic generative node embedding model을 적용했다. 구체적으로, 우리는 그래프가 주어질 시 graph topology와 node attribute에 가장 잘 fit하는 posterior distribution을 모델링하는 semi-implicit hierarchical construction을 혁신했다. SIVI와 함께한다면, posterior가 tractable하지 않아도, Monte Carlo estimation으로 posterior density를 evaluate할 수 있고, model flexibility와 expressive power가 높은 효율적인 model inference를 가능하게 한다. 우리의 semi-implicit graph variational auto-encoder (SIG-VAE)는 heavy tail, skewness, multimodality 등 posterior에는 드러나지만 현존하는 VGAE가 캡처하기 어려운 feature들을 모델링할 수 있다. 더 나아가서, Bernoulli-Poisson link function이 SIG-VAE의 decoder에 적용되어서 generative model의 flexibility를 늘리고, 대부분 sparse한 real-world network의 graph property를 더 잘 포착할 수 있다. SIG-VAE는 우리의 실험에서 검증한 여러 그래프 분석 task에서 end-to-end learning을 촉진했다. Link prediction에서, SIG-VAE는 SOTA한 method들을 상당히 크게 능가했다. Node classification과 graph clustering용으로 변형한 구조 또한 SOTA와 comparable했다. 더 나아가서 우리는 새로운 decoder가 real-world graph data의 statistics와 가까운 값을 가지는 sparse한 random graph를 generate할 수 있음을 보일 것이다. 이런 결과들은 SIG-VAE가 가지는 굉장한 가치를 보여준다.

1. Background
2. Variational graph auto-encoder (VGAE)

여러 node embedding method는 deterministic한 latent representation을 유도한다. Variational AutoEncoder (VAE) 개념을 graph에다가 확장함으로써, Kipf and Welling은 node를 latent space의 Gaussian random vector에다가 embed함으로써 다음의 문제를 해결하려고 했다.

* Problem 1

Graph 가 있고 adjacency matrix A와 M-dimensional node attribute 이 있다고 해 보자. 이 상황에서, node의 latent representation 의 probability distribution, 그러니까 를 구해보자.

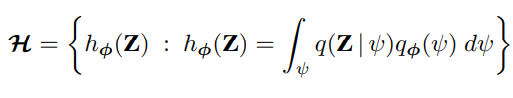
True posterior에 해당하는 는 대부분의 경우 intractable하다. Kipf와 Welling이 쓴 논문에서는, 이것은 다음의 Gaussian distribution으로 approximate된다.

,

여기서 는 node i에 대응하는 L-dimensional mean vector이고, 는 L-dimensional standard deviation vector이다. 의 parameter, 그러니까 은 두 개의 GCN을 이용해서 모델링되고 학습을 거치게 된다. 더 정확하게 말하자면, 들을 element로 가지는 matrix를 라고 하고 들을 element로 가지는 matrix를 라고 할 시 와 이 된다. 이렇게 Z가 주어지면, VGAE의 decoder는 로 쓸 수 있는 간단한 inner-product 기반 decoder이다. 이 모델의 parameter는 잘 알려진 ELBO function을 optimize함으로써 찾을 수 있다. 식은 가 된다. 여기서 은 와 동등한 것임을 주목해야 한다. 비록 VGAE는 promising하지만, variational inference에서 발생하는 잘 알려진 문제점은 posterior의 variance를 underestimate하는 것이다. 이런 현상이 발생하는 이유는 q의 제한 범위를 정하는 variational family의 representation power와, true posterior의 complexity 사이의 mismatch라고 할 수 있다. 게다가 asymmetric한 KL divergence 항의 이용으로 인해서 이 문제는 더 커진다.

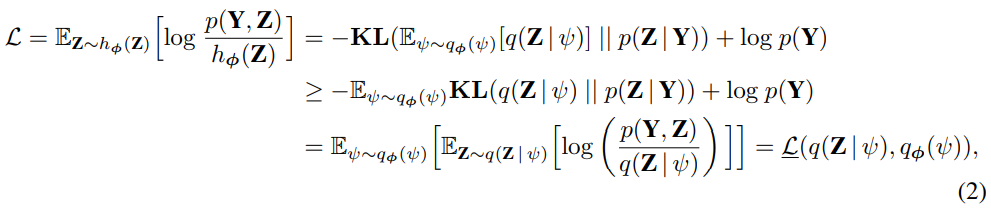
1. Semi-implicit Variational Inference (SIVI)

간단한 형태의 optimization을 유지하면서 posterior를 잘 characterize하기 위해서, semi-implicit variational inference (SIVI)가 Yin and Zhou에 의해 제기되었다. 이것은 또한 hierarchical variational inference와 auxiliary deep generative model과도 관련되어 있다. SIVI가 multimodal 또는 skewed distribution과 같은 복잡한 posterior distribution을 학습할 수 있다는 것이 드러났다(vanilla variational inference를 이용하면, prior와 latent space 내부의 posterior에 가해진 exponential family assumption으로 인해서 제대로 포착하지 못한다). SIVI는 posterior의 parameter 가 analytic하다기 보다는 implicit한 distribution에서 유도되었다고 가정한다. 이런 hierarchical construction은 유연한 mixture modeling을 가능하게 하며, model inference를 위한 간단한 형태의 optimization을 유지하면서 더 복잡한 posterior를 가능하게 한다. 더 구체적으로 말하자면, 이고 로 나타낼 수 있다(여기서 는 infer되는 distribution parameter를 나타낸다). 에 대해 marginalize[[2]](#footnote-2)할 시 variational parameter 로 index된 distribution family[[3]](#footnote-3) H로부터 유도된 random variable Z가 나오게 된다. 식으로 쓰면 다음과 같다.



Semi-implicit formulation에서 중요한 점은, original posterior 는 explicit하고 analytic하지만, marginal distribution인 는 보통은 implicit하다는 점이다[[4]](#footnote-4). 주목할 점은, 만약에 가 delta function[[5]](#footnote-5)과 같다면, 는 explicit한 distribution을 가진다는 점이다. 보통 variational inference는 latent dimension이 independent한 것을 가정하지만, semi-implicit variational inference는 그런 제한을 상정하지 않는다. 이것은 semi-implicit variational distribution이 매우 복잡한 multivariate distribution을 모델링할 수 있도록 한다.

Marginal probability density function 이 보통은 intractable하기 때문에, SIVI는 variational parameter를 optimize하기 위해서 다음의 ELBO로 lower bound를 설정한다.



여기서 Y는 observation을 의미한다. 가 을 유도하는 데 이용되었다. 이 lower bound를 optimize하는 것은 mixing distribution 을 point mass density로 유도할 수 있다. 이런 degeneracy issue를 해결하기 위해서, SIVI는 nonnegative regularization term을 추가해서, asymptotically하게 정확한 surrogate ELBO로 유도한다. Supplementary material에서 이거는 좀 더 다룰 것이다.

1. Normalizing flow (NF)

NF 또한 posterior distribution families를 enrich한다. SIVI와 비교 시, NF는 hierarchy 안의 mixing distribution에 대해 explicit한 density function을 할당하지만, SIVI는 가 reparameterizable하기만 하면 된다. 따라서 SIVI가 더 유연한 셈인데, SIVI posterior는 random noise를 NN과 같은 flexible한 함수를 이용해 transform함으로써 생성될 수 있기 때문에 graph analytics에서 이용될 때 더욱 유연해진다.

1. Baselines : Variational Inference with VGAE

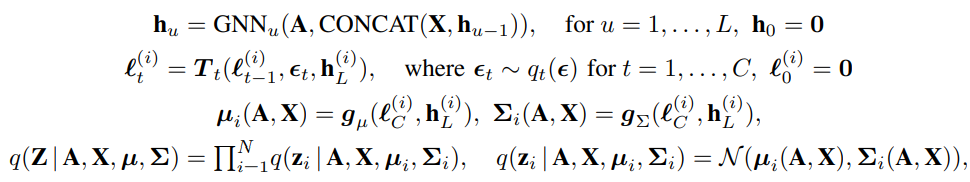
우리의 semi-implicit graph variational auto-encoder (SIG-VAE)를 제시하기 전에, SIVI와 NF를 VGAE와 다이렉트하게 결합하는 두 가지 baseline method를 소개할 것이다.

1. SIVI-VGAE

위의 Problem 1을 해결하면서도 posterior를 잘 characterize하고 VGAE framework에서 flexibility modeling을 잘 수행하기 위해서, 한 가지 naïve한 방법은 다음의 hierarchical formulation을 따라서 VGAE의 latent variable을 모델링하는 데 SIVI의 semi-implicit variational distribution을 이용하는 것이다.



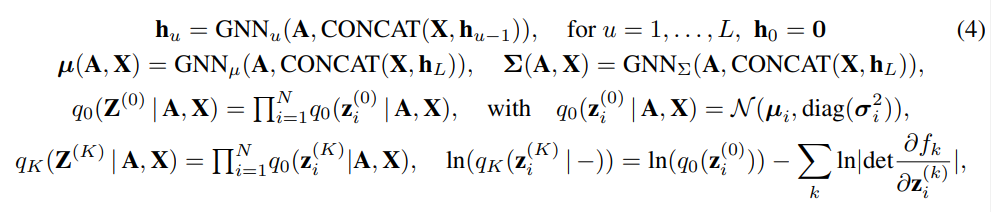
정확히 말하자면 reparameterizable한 로부터 sample될 수 있는, 로 parameterize될 수 있는 implicit prior distribution을 이용하는 것이다. 이런 hierarchical한 semi-implicit construct는 posterior의 mixture modeling을 더 유연하게 만들 뿐만 아니라 효율적인 model inference를 가능하게 한다(Ex. Deep NN으로 를 parameterize한다든지). 이 framework에서, GNN의 여러 layer에서 온 feature는, 각 node에서의 posterior distribution을 separate하게 derive하기 위해서 random noise와 concatenate한 다음에 여러 FC layer를 통해서 aggregate되고 transform될 수 있다. 더 구체적으로, SIVI-VGAE는 각 NODE를 대상으로 C개의 서로 다른 stochastic한 FC layer에다가 random noise를 독립적으로 집어넣는다. 식으로 쓰면 다음과 같다.



여기서 는 다 deterministic한 NN이고, i는 node index, L은 GNN layer의 수, 는 distribution 로부터 유도된 random noise이다. 여기서 주목할 점은, GNN은 GCN, GCN+Chebyshev filters, GraphSAGE, Jumping Knowledge (JK) networks, GIN 등 현존하는 graph neural network 중 어떤 것도 가능하다는 점이다. 의 output으로 이 나오면, 와 는 vanilla VAE를 따라 deterministic하게 결정되는 것이 아니라 하나의 RV가 된다. 하지만 이 방법으로는 implicit distribution을 만들었지만 이 distribution이 neighboring node 사이의 dependency를 제대로 포착하지 못할 수 있다. 주목할 점은, SIVI-VGAE를 우리의 SIG-VAE를 위한 naïve한 버전으로 만들었고(뒤에서는 Naïve SIG-VAE라고 부를 것이다), 이것은 network 안의 복잡한 dependency structure를 알아내기 위해 neighborhood sharing으로 구현된 구조이다. 또한 주목할 점은 SIVI의 첫 layer가 그냥 simple Gaussian이 아니라 NF와도 결합할 수 있다는 점인데, 이것은 나중의 연구가 할 일이다.

1. NF-VGAE

다른 variational inference를 이용해서 VGAE model의 flexibility를 향상시키는 것도 가능하다. NF를 이용하는 것이 한 가지 예시이다. 그러나, NF는 Jacobian을 계산하기 쉬운 deterministic transform function을 요구하고, 이것은 graph analytic task에서 마주하는 복잡한 dependency structure를 고려해보면 flexibility를 제한하는 꼴이다. 실제로 non-Gaussian VGAE, 그러니까 NF-based VGAE (NF-VGAE)를 다음의 식으로 구현해 보았다.

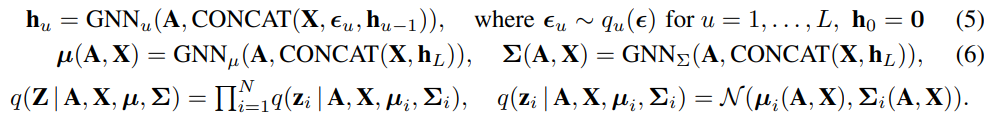


여기서 posterior distribution 는 distribution 를 따르는 Gaussian RV 을 K개의 invertible한 differentiable한 transformation 를 거쳐서 연속적으로 transform함으로써 얻을 수 있다. 이 내용은 supplementary material에서 좀 더 다룰 것이다. NF-VGAE는 2-step inference method인데, 1) Gaussian RVs로 시작하고 2) 일련의 invertible한 mapping을 이용해서 그 Gaussian RVs를 transform시키는 것이다. NF-VGAE에서, initial distribution 의 deterministic nature로 인해서 neighborhood distribution sharing 없이 GNN output layer가 deterministic하게 결정된다는 것을 다시 한 번 강조하고 싶다.

1. Semi-Implicit Graph Variational Auto-Encoder (SIG-VAE)

위의 두 baseline model은 더 flexible하고 복잡한 posterior를 approximate할 수 있지만, 이러한 trivial한 조합은 neighboring node 사이의 uncertainty를 propagate할 수 없기 때문에 graph dependency structure를 완전히 써먹지 못한다. Graph data를 다룰 시 복잡한 posterior를 알아내기 위해 필수적인 uncertainty propagation을 위해서 SIG-VAE를 고안했다. SIG-VAE는 variational inference와 VGAEs를 neighborhood sharing scheme을 이용해서 더 잘 통합할 것이다.

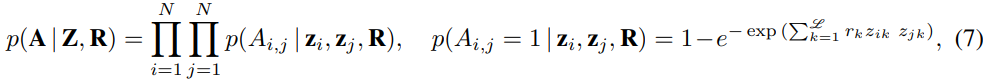
Tractable한 posterior inference를 위해서, SIG-VAE는 여러 stochastic layer의 hierarchy로 구성되어 있다. 구체적으로, 첫 stochastic layer 는 reparameterizable하고 analytic한 probability density function을 가지고 있다. 그 뒤에 추가되는 layer는 reparameterizble하고, layer로부터 효율적으로 sampling이 가능하다. 더 자세히 쓰자면, L개의 서로 다른 stochastic layer에다가 random noise를 넣는 hierarchical encoder를 SIG-VAE에 도입했다.



위의 equation에 있는 는 RV이고 그러므로 는 marginalization을 거친 후에 꼭 Gaussian인 것은 아니게 된다. 는 distribution 로부터 온 N-dimensional random noise가 되고, 는 로부터 유도된 sample의 type이 X와 같도록 선택된다 (ex. X가 categorical하가면, 로는 Bernoulli가 좋은 선택일 것이다. 이런 random noise와 node attribute를 concatenate함으로써, GNN의 결과는 deterministic vector가 아니라 random variable이 된다. 이 random variable의 expressive power는 SIG-VAE로 상속되어서, Gaussian, exponential family, von Mises-Fisher posterior distribution보다 복잡한 distribution도 latent representation이 나타낼 수 있게 한다.

SIG-VAE에서 각 node의 latent posterior를 infer할 시, neighboring node의 distribution을 incorporate함으로써, GNN으로부터 온 deterministic한 feature를 공유하는 방법보다 더 잘 graph dependency structure를 알아낸다. 구체적으로, stochastic layer u에서의 input은 CONCAT(X,)이기 때문에 그 뒤의 stochastic layer는 neighboring node로부터 온 정보를 종합해서 mixing distribution을 결과로써 내놓는다. (5)-(6)번 식에서 stochastic distribution parameter에 작동하는 SIG-VAE의 flexibility는 graph analytic task에서 neighborhood sharing이 더 높은 performance를 내놓도록 한다. SIG-VAE에서 equation (5-6)을 거치면서 GNN을 이용하고 input에다가 noise를 집어넣어서 나온 uncertainty propagation이 더 복잡한 posterior를 더 충실하게 encode하는 핵심이라고 생각한다. 주목할 점은 (5)번 식이 NF-VAE와는 다르다는 점이다(NF-VAE에서는 GNN output layer가 deterministic하다). 실험을 거치면서, 이런 uncertainty neighborhood sharing이 SIG-VAE가 훌륭한 graph analysis performance를 보이는 핵심이라는 점을 보일 것이다.

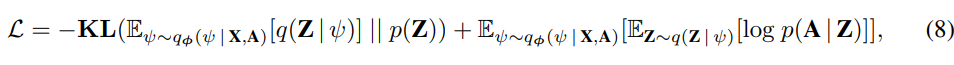
더 나아가서, variational inference의 flexibility를 늘리는 것이 real-world graph data를 더 잘 모델링하는 데에는 충분하지 않은데, 이것은 generative model의 optimal solution이 변화하지 않았기 때문이다. SIG-VAE에서는, Bernoulli-Poisson link를 decoder에 도입해 generative model의 expressiveness를 더욱 늘렸다. 다른 decoder의 potential extension은 필요하다면 SIG-VAE와 융합될 수 있다. 라고 하자. 그러면 다음의 식으로 decoder를 쓸 수 있다.[[6]](#footnote-6)



여기서 은 diagonal element로 를 가지는 diagonal matrix이다.

* 1. Inference

SIG-VAE에서 model inference를 위한 ELBO를 유도하기 위해서는, 가 어떠한 distribution에서 온다는 것을 고려해야 한다. 그러므로, ELBO는 다음의 식으로 나타낼 수 있다.



여기서 는 위의 (1)에서 정의된다. Marginal probability density function 는 보통 intractable하니, ELBO L의 Monte Carlo estimation은 금지된다. 이 문제를 해결하고 SIG-VAE의 variational parameter를 추론하기 위해, 다음의 식으로 ELBO의 lower bound를 설정할 수 있다.



자세한 implementation detail과 surrogate ELBO의 유도는 supplementary material에서 찾아볼 수 있다.

1. Experiments

SIG-VAE를 다음의 graph analytic task에 대해서 performance를 측정해 보았다. 1) VGAE와 비교시 SIG-VAE의 interpretability 2) node attribute를 포함하는 / 포함하지 않는 real-world graph dataset에서의 link prediction 3) graph generation 4) label이 있는 citation graph에서의 node classification. 이 모든 실험에서 GNN module로는 GCN이 쓰였고, SIG-VAE, Naïve SIG-VAE, NF-VGAE를 Tensorflow로 구현했다. PyGSP package가 synthetic graph를 generate하는 데 이용되었다. Supplementary material에서 실험과 graph data statistics에 대한 더 자세한 정보를 찾아볼 수 있다.

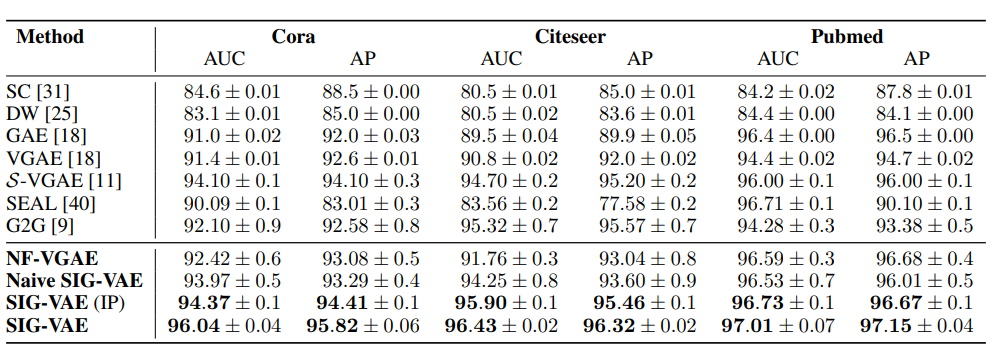
* 1. Interpretable latent representations

우선 node latent representation의 approximated variational distribution을 묘사함으로써 SIG-VAE의 expressiveness를 나타낼 것이다. SIG-VAE는 200 node와 1244개 edge로 만들어진 Swiss roll graph에서 VGAE보다 graph structure를 더 잘 학습했고 더 interpretable한 embedding을 가지고 있었다. 공정한 비교를 위해서, VGAE와 SIG-VAE 모두 inner-product decoder와 같은 수의 parameter를 가지도록 동일하게 구현했다. Node attribute로는 identity matrix 을 이용했고 latent space dimension은 3이 되도록 설정했다. 이 그래프는 평면이랑 비슷한 간단한 구조이다. Inner-product decoder는 latent vector 사이의 각에다가 정보가 embed된다고 가정하기 때문에, latent space에다가 Swiss roll graph의 node들을 curve 형태로 embed할 것이라고 예상했다. Figure 2에서 볼 수 있듯이, SIG-VAE가 VGAE보다 더 interpretable한 planar latent structure를 유도해냈다. 또한 Figure 3에다가 랜덤으로 선택한 node 5개의 posterior distribution을 그려 놓았다. 추론된 distribution은 multi-modal하거나, skewed이거나, non-symmetric하거나, 날카롭고 급격한 변화를 가지고 있을 수 있다. 이런 복잡한 distribution은 모델로 하여금 intrinsic한 graph structure를 포착할 수 있는 더 실제 같은 embedding을 생성할 수 있도록 돕는다. 왜 multi-modality가 등장하는지를 설명하기 위해서, Asynchronous Fluid를 이용해 우리가 알아낸 community를 서로 다른 색으로 Figure 4에다가 나타냄으로써 Swiss Roll graph를 visualize해 보았다. Graph 안의 community를 더 잘 visualize하기 위해서 Figure 2(a)에 나온 layout과는 다른 layout을 이용했다. Figure 3에 있는 빨간색 node 3개와 주황색 node 3개는 multi-modal (skewed) distribution을 가지는 node들이다. 이런 multi-modal posterior는 서로 다른 community에 존재한다; 그러므로, 확률적으로는 그 node들은 multiple community에 소속될 수 있는 것이다. Supplementary material을 보면 torus graph를 가지고 진행한 추가적인 결과와 discussion이 나와있다.

* 1. Accurate link prediction

Real-world graph dataset을 가지고 link prediction 실험을 진행했다. 결과를 보면, SIG-VAE가 모든 benchmark dataset에서 잘 알려진 baseline method들과 SOTA method를 크게 능가했다. 두 종류의 dataset을 생각할 수 있었다 : node attribute가 있는 경우와 없는 경우. Kipf와 Welling의 연구에서 했던 것처럼, validation set이 network link의 5%를 포함하게 하고 test set이 10%를 포함하도록 dataset을 split했다. Model parameter를 learning rate가 0.0005인 채로 3500 epoch 동안 학습을 진행했고, early stopping을 위해 validation set을 이용했다. Latent space dimension은 16으로 설정했다. SIG-VAE, Naive SIG-VAE, NF-VGAE의 hyperparameter는 모든 dataset에 대해 같았다. 공정한 비교를 위해서, 위 세 가지 method는 default VGAE와 유사한 수의 parameter를 가진다. Supplementary material에 더 자세한 implementation detail이 나와 있다. 이 그래프의 test set에서 사전에 link들을 좀 제거해 놨고, 이 test set을 가지고 average precision (AP)와 area under the ROC curve (AUC)를 이용해 10 run 동안의 performance를 계산했다.

1. With node attributes

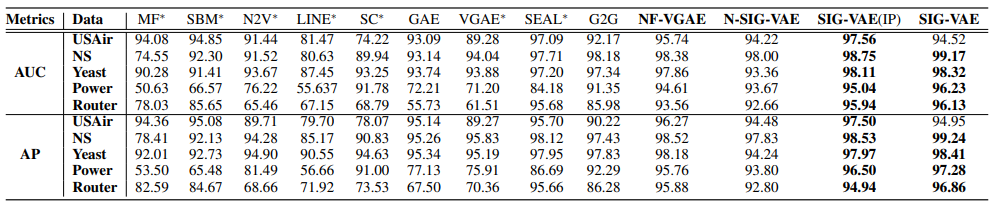


*Table 1 : node attribute를 가지는 network에서의 link prediction performance*

Node attribute를 가지는 3개의 graph dataset을 이용했다(Cora, Citeseer, Pubmed). Node attribute의 수는 각각 1433, 3703, 500개였다. 다른 dataset에 대한 정보는 supplement Table 1에 요약되어 있다. SIG-VAE, Naïve SIG-VAE, NF-VGAE의 결과를 6개의 SOTA method (Spectral Clustering, DeepWalk, GAE, VGAE, S-VGAE, SEAL)과 비교했다. Encoder에서 semi-implicit hierarchical variational distribution의 이점을 분명하게 나타내기 위해서, SIG-VAE에 inner-product decoder를 이용했다.

[40, 18, 11]에 나타나 있는 대로 competing method의 hyperparameter를 설정했다. Table 1에서 볼 수 있듯이, AUC와 AP에서 모두 SIG-VAE가 SOTA method보다 상당히 나은 performance를 보여주었다. SIG-VAE의 standard deviation 또한 다른 method보다 작았는데, 이것은 semi-implicit variational inference가 안정적임을 나타낸다. Baseline VGAE와 비교할 시, SIGVAE (inner-product decoder와 Bernoulli-Poisson link decoder에서 모두), Naïve SIG-VAE, NF-VGAE가 가지는 더 flexible한 posterior가 link prediction accuracy를 분명히 늘렸다. 이것은 이런 graph structured data에 대해서는 Gaussian assumption이 맞지 않았다는 것을 의미한다. Inner-Product decoder (IP)를 가지는 SIG-VAE가 Naïve SIG-VAE와 NF-VGAE와 비교할 시 보여주는 performance 향상은 neighboring node sharing의 이점을 분명하게 보여준다(특히 작은 그래프에서 그렇다). 큰 graph인 Pubmed에서도(VGAE가 S-VGAE와 performance가 비슷했다), SIG-VAE가 가장 높은 link prediction accuracy를 보여주었고, non-Gaussian prior, neighborhood distribution, sparse Bernoulli-Poisson link decoder와 같은 구성요소들이 전부 중요하다는 것을 보여주었다.

1. Without node attributes



*Table 2 : node attribute가 없는 network를 가지고 진행한 link prediction의 AUC와 AP. \*는 코뭏과 Chen의 연구에서 보고된 값이다.*

이제 node attribute를 가지는 5개의 graph dataset (USAir, NS, Router, Power, Yeast)를 가지고 실험을 진행했다. Supplement Table 1에다 data statistics가 나와 있다. 우리의 모델의 performance를 7개의 SOTA method (Matrix Factorization, Stochastic Block Model, node2vec, LINE, Spectral Clustering, VGAE, S-VGAE, SEAL)과 비교해 보았다.

Baseline method에서는, Zhang et al.이 이용한 값을 이용해 hyperparameter를 설정했다. Node attribute를 가지지 않는 dataset이다 보니까, SIG-VAE에서 2-stage learning process를 이용했다. 우선, system에다가 5-dimensional Bernoulli noise를 주입하면서 각 node embedding이 128-dimensional latent space에 생성된다. 그 뒤 학습한 embedding이 second stage의 node feature로 이용되어, 64-dimensional noise를 SIG-VAE에 주입하는 동시에 16-dimensional embedding을 학습한다. Empirical한 실험을 거치면서, 이 two-stage learning이 end-to-end learning보다 더 빠르게 converge한다는 것을 알아냈다. Naïve SIG-VAE와 NF-VGAE에서도 같은 절차를 거쳤다.

Table 2에서 볼 수 있듯이, SIG-VAE는 AUC와 AP에서 모두 다른 method보다, 특히 baseline VGAE와 비교했을 시 꾸준하게 performance가 좋았다. 흥미로운 점은, 우리가 제시한 Bernoulli-Poisson decoder가 NS와 Router dataset과 같은 sparse한 graph에 대해서는 performance가 좋았는데, inner-product decoder를 장착했을 시 상대적으로 더 dense한 USAir dataset에서의 performance가 더 좋았다. Baseline VGAE와 비교할 시, Naïve SIG-VAE와 NF-VGAE가 모두 AUC와 AP 측면에서 실험 결과가 크게 좋았고, 더 flexible한 posterior의 이점을 보여주었다. SIG-VAE를 Naïve SIG-VAE와 NF-VGAE 이렇게 두 개의 flexible inference method와 비교해보면, SIG-VAE가 graph topology와 관련된 flexible한 posterior를 모델링할 수 있다는 것뿐만 아니라, SIG-VAE가 Gaussian assumption에 제한되지 않는다는 것을 보여준다. 그리고 이 Gaussian assumption은 inner-product decoder와는 잘 맞지 않는 점이다. Power graph의 link prediction 결과는 이 사실을 잘 보여주는데, SIG-VAE가 VGAE와 비교할 시 accuracy를 34%나 늘렸다. Supplementary material을 보면 서로 다른 회차의 실험에서 stdev를 보여주는데, 이것이 다시 SIG-VAE의 stability를 보여준다.

Node attribute를 이용하지 않는 citation graph에서의 link prediction task가 주어졌을 시, inner-product decoder를 가지는 SIG-VAE를 evaluate하기 위해서 ablation study가 행해졌다. [AUC, AP]는 Cora에서 [91.14, 90.99]였고, Citeseer에서 [88.72, 88.24]였다. 이 값은 Table 1에 있는 attribute를 가지는 SIG-VAE에서 보고된 결과보다는 낮지만, 다른 method들 (node attribute를 가지는 경우를 포함해서)과 비교할 시 competitive했으며, 이것은 graph structure를 SIG-VAE가 이용할 수 있는 능력을 보여준다. SEAL과 같은 다른 method는 node attribute가 없는 graph를 잘 다뤘고, VGAE와 같은 다른 method들은 node attribute가 있는 graph를 잘 다뤘지만, SIG-VAE는 attribute가 있건 없건 간에 performance가 좋았다. 이것은 SIG-VAE가 node의 implicit distribution을 학습할 수 있고, 이 점이 node attribute가 없이도 graph structure를 알아내는 데 굉장히 강력한 점이기 때문이다.

* 1. Graph generation

Generative model로서 SIG-VAE의 flexibility를 설명하기 위해서, inferred embedding representation으로부터 새로운 graph를 generate해 보았다. 예를 들어서, Cora dataset을 가지고 진행해 본다면 이 dataset의 density는 0.00143이고 average clustering coefficient는 0.24인데, SIG-VAE는 Cora에 맞는 새로운 network parameter를 추론한다. 이렇게 추론한 inferred posterior와 학습을 거친 decoder를 가지고, decoder의 diagonal matrix R의 diagonal element인 에 맞게 corresponding graph가 만들어졌고, 그 graph의 graph statistics가 original graph에 가까운지를 체크했다. 잠시 주목할 점은, inferred 를 0.01보다 작은 값으로 shrunk했다는 점이다. 이렇게 SIG-VAE를 가지고 generate된 graph의 density는 0.00147이었고 average clustering coefficient는 0.25였으며, 이 값은 original graph의 값과 굉장히 유사하다. 또한 inner-product decoder+SIG-VAE 모델과 VGAE를 가지고 새 graph를 generate해 보았다. 이 경우 SIG-VAE(IP)와 VGAE가 generate한 graph의 density와 average clustering coefficient는 0.1178과 0.49였고, 이것은 sparse graph에 대해서는 inner-product decoder가 좋지 않은 선택지임을 보여준다.

* 1. Node classification & graph clustering

또한 loss function이 graph reconstruction term과 semi-supervised classification term을 포함하도록 변형함으로써 label이 있는 citation graph에 대해 SIG-VAE로 node classification을 수행해 보았다. 결과는 Table 3에 나와 있다. 우리의 모델은 강한 generalization property를 보여 주었는데, SOTA method와 비교 시 이 node classification을 위해 train되지 않았음에도 불구하고 competitive한 결과를 보여주었다. SIG-VAE가 missing edge에 대해 가지는 robustness를 보여주기 위해서, 10, 20, 50, 70%의 edge를 node attribute를 그대로 둔 채로 제거했다. Cora dataset을 가지고 10번 진행한 mean accuracy는 각각 79.5, 78.7, 75.3, 60.6이었다.

SIG-VAE는 link prediction에서 SOTA한 performance를 보여주었고, 다른 task에서 comparable한 결과를 낳았다. 이 결과는 분명하게 SIG-VAE가 여러 graph analytic task를 수행하는 데 가지는 잠재력을 보여준다.

1. Conclusion

Semi-implicit hierarchical variational distribution과, Bernoulli-Poisson link decoder를 가지는 VGAE의 장점을 결합함으로써, SIG-VAE는 graph가 주어질 시 node embedding의 posterior distribution을 나타내는 representation power를 강화했고, 그럼으로써 latent space에서 graph structural information과 node attribute information을 잘 capture할 수 있었다. Asymptotically하게 정확한 surrogate ELBO를 제시함으로써, SIG-VAE model inference의 optimization은 SGD를 통해서 가능해졌고, variational distribution의 flexibility를 해치지 않았다. 서로 다른 graph dataset을 가지고 진행한 우리의 실험은 여러 graph analysis application에서 interpretable한 latent representation을 SIG-VAE가 다룰 수 있는 잠재력을 보여 주었고, 이것은 주어진 graph에서 neighborhood node의 distribution을 diffuse하는 hierarchical construction 덕분이었다.

1. Mean-field approximation이란, Bayesian posterior inference에서 우리가 알아야 하는 bayesian 항에서 분모 p(x)를 해결하기 위해서 variational inference를 적용하는 것을 말한다. Posterior density가 라고 할 시, 이걸 새로운 parameter에 대한 density인 라고 쓰는 것이다. 이렇게 bayesian posterior density를 새로운 density로 근사하는 것을 mean-field approximation이라고 부른다. [↑](#footnote-ref-1)
2. 특정한 parameter에 대해 marginalize한다는 것은, 그 parameter에 대한 확률과 그 parameter가 주어질 시 다른 확률변수의 확률을 곱해서 적분하는 것을 의미한다. 예를 들어서, P(기분|날씨,커피)=P(기분|날씨,커피=아아)+P(기분|날씨,커피=라떼)+P(기분|날씨,커피=바닐라라떼)+… 와 같은 것을 의미한다. 실제로는 적분을 이용한다. [↑](#footnote-ref-2)
3. Distribution family H가 variational parameter 로 index된다는 말은, 가 가지는 값에 따라서 달라지는 들의 집합이 H가 된다는 것을 의미한다. [↑](#footnote-ref-3)
4. 여기서 Explicit과 Implicit의 의미를 다시 한 번 짚고 넘어가자. Explicit probabilistic model은 우리가 관찰하고 있는 random variable x가 있다고 할 시, 이 x를 explicit하게 parameterize해서 라는 log likelihood function을 따른다고 하는 것이다. ML에서 대부분의 모델은 explicit하다. 반면에 implicit한 모델은 data generation을 가능하게 하는 stochastic한 procedure를 정의하는 식이다. [↑](#footnote-ref-4)
5. Delta function은 에서 의 값을 가지고 그 외 0을 가지는 함수를 의미한다. [↑](#footnote-ref-5)
6. Delta function을 1변수 크로네커-델타로 생각해 보면, input이 0일 때 output이 1이고 그 외 input이 들어올 시 output이 0인 함수로 생각할 수 있다. 가 Poisson을 따르니까 자연스럽게 0 이상이 되고, 가 0보다 클 시 delta는 0이 된다. 따라서 이려면 Poisson이 0이어야 하고, Poisson이 0일 확률은 Poisson RV의 식을 보면 k=0일 때이니 확률은 가 된다. 그래서 (7)이 유도된다. [↑](#footnote-ref-6)