Variational Graph Auto-Encoders

VGAE变分图自动编码器，主要为了在图中寻找合适的Embedding向量，并通过Embedding向量实现**图重构**，获取到的Embedding可以支撑下游任务。

自编码器

自编码器是通过减少隐藏层神经元个数实现重构样本，自编码器为了尽可能的复现输入的数据，其隐藏层必须捕捉输入数据的主要特征，从而找到代表原输入数据的主要成分。
VAE是变分贝叶斯和神经网络的结合

VAE假设先验$p(z)$和近似后验$q(z \mid x)$都遵循高斯分布,即令$p(z) \sim N(0,1)$

假设$z \sim N(\mu, \sigma 2), \mathrm{z}=\mu+\sigma \varepsilon$, 其中$\varepsilon \sim N(0,1)$。$\mu$ 和σ均由神经网络参数化。基于计算出的潜在变量z，利用解码器网络来重构输入数据。

变分贝叶斯

统计模型由观察变量x未知参数θ和隐变量z组成，生成模型是通过隐变量来估计观察变量：
$p_\theta(z) p_\theta(x \mid z)$

但很多情况下，后验概率并不容易得到，所以通过其他方式来近似估计这个后验概率。

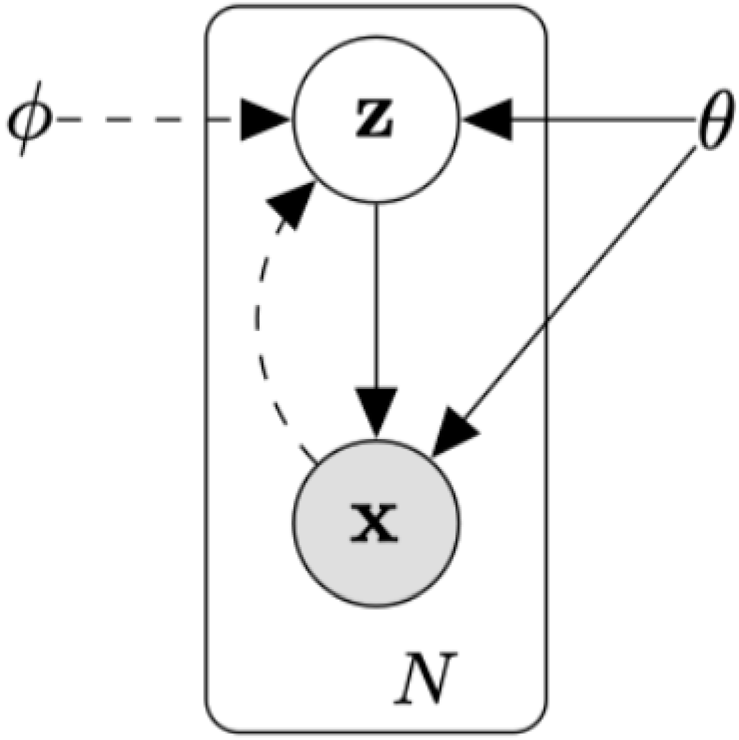
传统贝叶斯是采用MCMC采样方法，通过抽取大量的样本后给出后验概率的近似值，计算效率低

变分贝叶斯是把原本的统计推断问题转换为求两个分布最优距离的问题，利用一种分析方法近似隐变量的后验分布

VAE是用神经网络来学习变分推导的参数，从而得到后验推理的似然估计

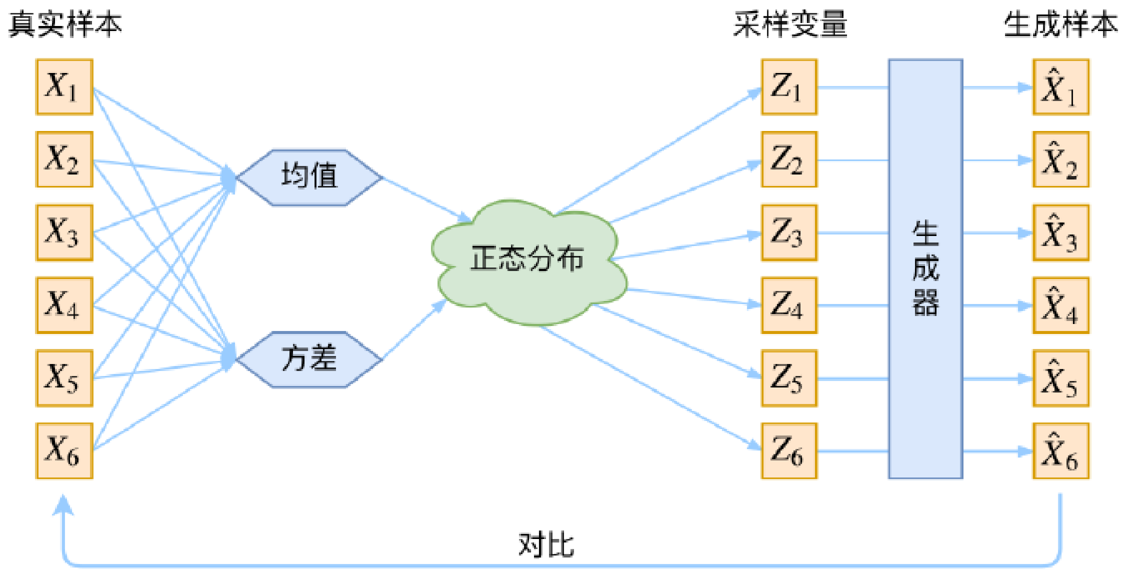
实线表示贝叶斯推断统计的生成模型$p_\theta(z) p_\theta(x \mid z)$（后验概率）

虚线表示变分近似$q_\phi(z \mid x)$



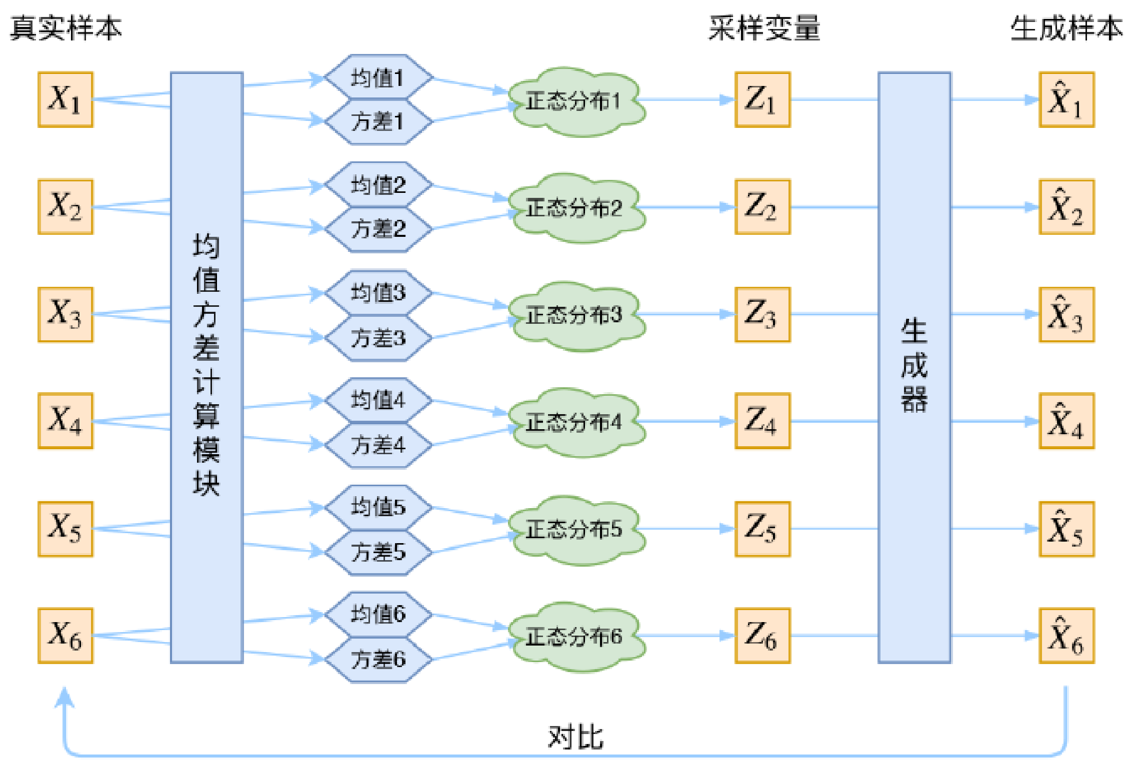
提出了AEVB算法来让$q_\phi(z \mid x)$近似$p_\theta(x \mid z)$同时把最大似然函数的下界作为目标函数，从而避开了后验概率计算，并将问题转化为最优化问题，并可以用随机梯度下降来进行从参数优化。

VAE 模型中，我们假设 $\boldsymbol{q}_\phi(\boldsymbol{z} \mid \boldsymbol{x})$ 这个后验分布服从正态分布，并且对于不同样本来说都是独立的，即样本的后验分布是独立同分布的。



这样的结构无法保证通过学到的分布进行采样得到的隐变量 z_i 能够与真实样本 x_i 一一对应，所以就无法保证学习效果了。

所以 VAE 的每个样本都有自己的专属正态分布：



这样，我们便能通过每个样本的专属分布来还原出真实样本。

这也是论文中最重要的一点：

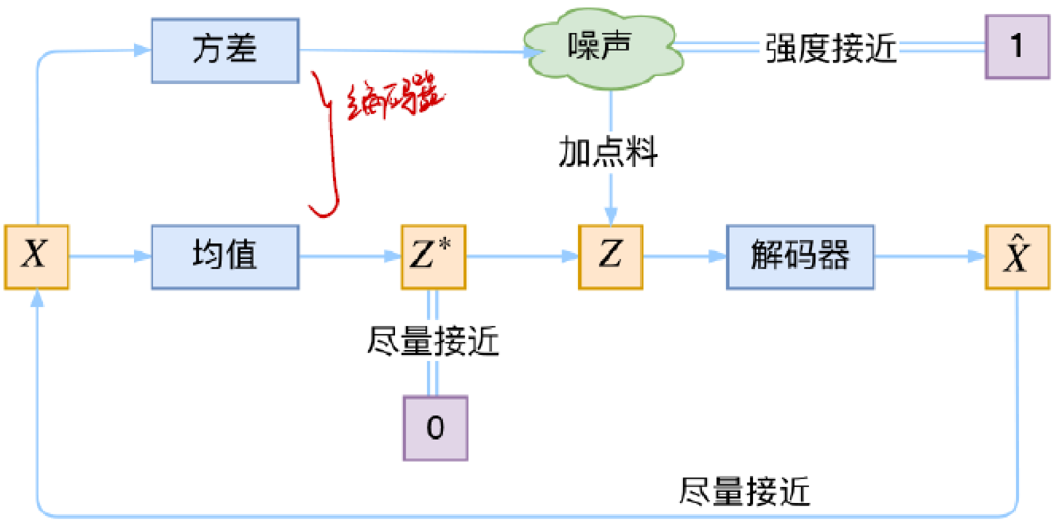
$$\log q_\phi(\mathbf{z}|\mathbf{x}^{(i)}) = \log N(\mathbf{z};\mu^{(i)},\sigma^{2(i)},\mathbf{I})$$

VAE通过构建两个神经网网络来分别学习均值和方差$\mu_k = f_1(\mathbf{X}_k), \log \sigma_k^2 = f_2(\mathbf{X}_k)$这样就可以得到样本$X_k$的专属均值和方差了，然后从专属分布中采样出$Z_k$然后通过生成器得到$\hat{\mathbf{x}}_k = g(\mathbf{Z}_k)$通过最小化重构误差来进行约束$D\left(\hat{\mathbf{X}}_k, \mathbf{X}_k\right)$

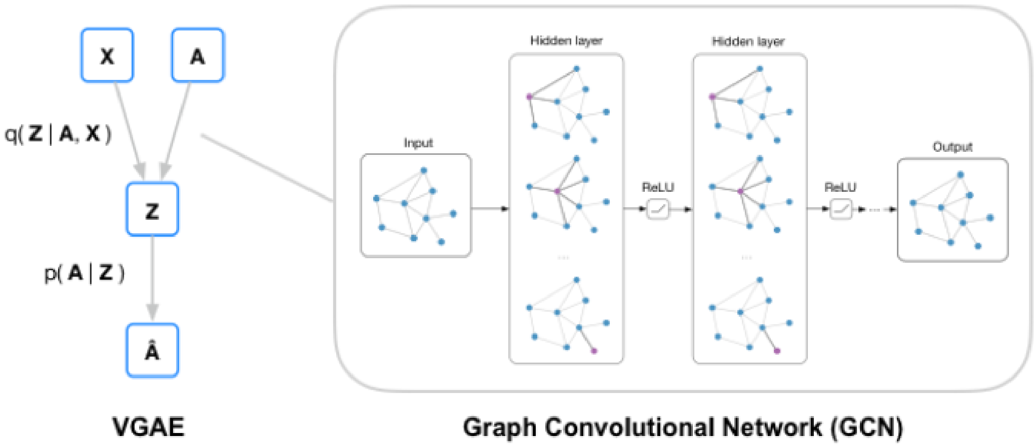
但隐变量是通过采样得到的，而不是经过编码器算出来的。这样的重构过程中免不了受到噪声的影响，噪声会增加重构的难度，不过好在这个噪声的强度可以通过方差反应，方差可以通过一个神经网络得到计算，所以最终模型为了更好的重构会尽量让模型的方差为零，而方差为零时，就不存在随机性了，这样模型最终会是一组均值，便退化成了普通的 AutoEncoder

为了防止噪声不起作用，VAE会让所有后验分布都向标准正太分布看齐（KL散度计算）

> 变分自编码中的变分是指变分法，用于对泛函 $KL(p|q)$ 求极值。
>
> 我们将约束两个分布的 KL 散度加入到损失函数中，则有：
>
> KL散度
>
> $$L = D(\hat{\mathbf{X}}_k, \mathbf{X}_k) + \frac{1}{2}\sum_{i=1}^{d}(\mu_i^2 + \sigma_i^2 - log\ \sigma_i^2 - 1)$$
>
> 简单来说，VAE 的本质就是利用两个编码器分别计算均值和方差，然后利用解码器来重构真实样本，模型结构大致如下：



VGAE



其中，X为节点的特征矩阵，A为邻接矩阵，先利用后验概率得到隐变量 Z，再用隐变量重构邻接矩阵 A。

VGAE 的编码器是一个两层的图卷积网络：

$$q(\mathbf{Z}|\mathbf{X},\mathbf{A}) = \Pi_{i=1}^{N} q(z_i|\mathbf{X},\mathbf{A})$$

其中，后验概率和 VAE 的解决方案一致：

$$q(\mathbf{z}_i|\mathbf{X},\mathbf{A}) = N(\mathbf{z}_i|\mu_i, diag(\sigma_i^2))$$

其中，$\mu = GCN_\mu(\mathbf{X},\mathbf{A})$ 是特征向量的均值；$log\,\sigma = GCN_\sigma(X,A)$ 是节点向量的方差。

两层卷积神经网络定义为：

$$GCN(\mathbf{X},\mathbf{A}) = \tilde{\mathbf{A}}\, ReLU(\tilde{\mathbf{A}}\mathbf{X}\mathbf{W_0})\mathbf{W_1}$$

其中，$GCN_\mu(\mathbf{X},\mathbf{A})$ 和 $GCN_\sigma(\mathbf{X},\mathbf{A})$ 共享第一层参数 W_0，不共享第二层参数 W_1；$\tilde{\mathbf{A}} = \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$ 是对称标准化邻接矩阵。

VGAE 的解码器则是利用隐变量的内积来重构邻接矩阵：

$$p(\mathbf{A} \mid \mathbf{Z}) = \prod_{i=1}^{N}\prod_{j=1}^{N} p(A_{ij}|\mathbf{z}_i,\mathbf{z}_j)$$

其中，$p(A_{ij} = 1|\mathbf{z}_i,\mathbf{z}_j) = \sigma(\mathbf{z}_i^T\mathbf{z}_j)$.

损失函数也是包括两部分：

$$L = \mathbb{E}_{q(\mathbf{Z}|\mathbf{X},\mathbf{A}))}[log\, p(\mathbf{A} \mid \mathbf{Z})] - KL[q(\mathbf{Z} \mid \mathbf{X},\mathbf{A})||p(\mathbf{Z})]$$

其中，$p(\mathbf{Z}) = \Pi_i N(\mathbf{z}_i|0,\mathbf{I})$。