

该论文和之前的 ConvKB 是同一个作者，可以看做是 ConvKB 的一个扩展，对于一个三元组 (s, r, o)，其中 s 为头实体，o 是尾实体，r 为头尾实体之间的关系。CapsE 首先学习一个 3 列的矩阵，每一列分别表示头实体、关系和尾实体的向量，之后使用多个卷积核对该矩阵进行卷积操作，得到相应的多个特征向量。到目前为止，CapsE 和 ConvKB 是一致的，都是使用了 CNN 来提取三元组内部的深层关系。

在本篇论文中，胶囊网络被加在了特征向量的后面，特征向量变换之后便成为了第一层的胶囊网络，每一个向量为一个胶囊，即下图中的 u1、u2、u3 和 u4。在经过路由过程 (routing precessing) 和 squash 激活函数之后，信息被传递到了第二层的胶囊网络。第二层胶囊网络中仅有一个胶囊，该胶囊对应的向量的模长即为 (s, r, o) 三元组存在于知识图谱中的概率。

Motivation:

隐层表示的可分解性，以接受直觉的解释。

VAE:

VAE 是一个概率模型 (probabilistic model)，它把深度模型和概率方法结合了起来，也因此取名 Variational Autoencoder。它的主要作用是生成新的相似的数据而非学习一个低维的隐性表达。

Let's label this distribution as  $p^\theta$ , parameterized by  $\theta$ . The relationship between the data input  $x$  and the latent encoding vector  $z$  can be fully defined by:

VAE takes the inference model  $q_\phi(z|x)$  to be an encoder and the likelihood model  $p^\theta(x|z)$  to be a decoder. Here  $\theta$  and  $\phi$  are neural network parameters,

VAE 是为每个样本构造专属的正态分布，然后采样来重构。

VAE 的关键是在隐层的处理，引进了  $Q(Z|X)$  来近似  $P(Z|X)$ ，这样就可以通过采样  $Z$  来最大化  $P(X)$ ， $P(X)$  是网络的逻辑输出。是需要最大化的目标，也就是最大似然。

$\beta$ -VAE 的目标是学习独立的特征，让某种特征对应某个生成因素，而独立于其他因素。

我们试想在  $z$  处加一个“瓶颈” $\beta$ ，它像一个沙漏，进入的信息量可能有很多，但是出口就只有  $\beta$  那么大，所以这个瓶颈的作

用是：不允许流过  $z$  的信息量多于  $\beta$ 。跟沙漏不同的是，沙漏的沙过了瓶颈就完事了，而信息过了信息瓶颈后，还需要完成它要完成的任务（分类、回归等），所以模型迫不得已，只好想办法让最重要的信息通过瓶颈。这就是信息瓶颈的原理！

Beta 越大，overlap 越大。Beta 为 0 的时候，类似于一个 Look-up table，就几乎没有 overlap。When  $\beta > 1$ , it applies a stronger constraint on the latent bottleneck and limits the representation capacity of  $z$ . For some conditionally independent generative factors, keeping them disentangled is the most efficient representation. Therefore a higher  $\beta$  encourages more efficient latent encoding and further encourages the disentanglement. Meanwhile, a higher  $\beta$  may create a trade-off between reconstruction quality and the extent of disentanglement. 把十字换成正方形，没有太多的耦合，（耦合举例，可以看成是多个高斯分布）当 beta 很大的时候，相当于 latent factor 内部的变量整齐的排列，很少的 overlap，并不是解耦合的定义。

各向同性 Isotropic: 指物体的物理、化学等方面的性质不会因方向的不同而有所变化的特性，即某一物体在不同的方向所测得的性能数值完全相同，亦称均质性。

各向异性 Anisotropic: 指物质的全部或部分化学、物理等性质随着方向的改变而有所变化，在不同的方向上呈现出差异的性质。beta 越大，后验和先验越像， $z$  会聚成几个团，失去了对数据的区分或者准确的建模，