

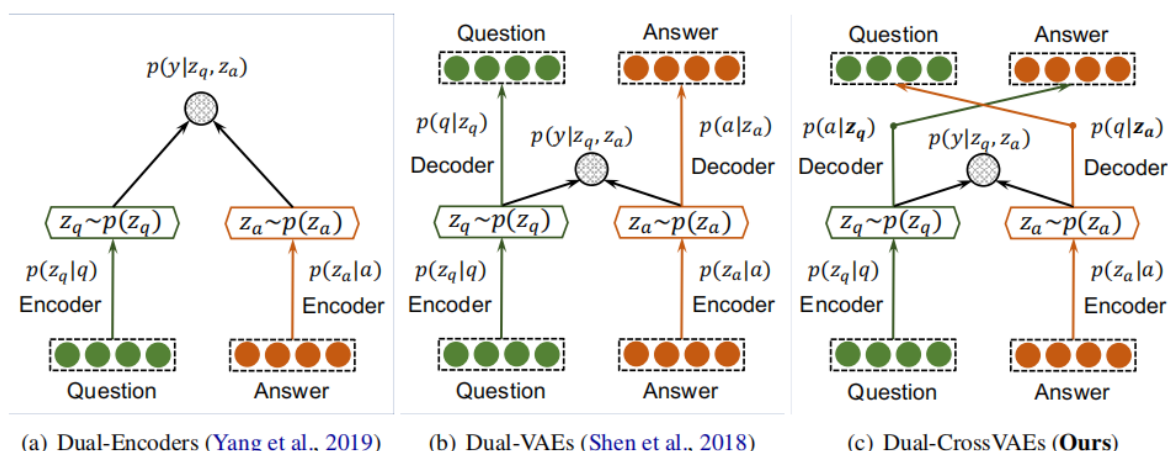
Crossing Variational Autoencoders for Answer Retrieval

Conference: 2020 ACL

arxiv:

idea:VAE

Model Arcitecture



- 图(a)为传统的孪生网络结构，也称为‘Dual Encoder’结构，即双Encoder结构，使用两个**独立**的编码器对问题和答案进行分别编码
- 图(b)为双VAE结构，即‘Dual-VAEs’结构，通过两个**独立**的VAE分别得到隐状态 z_q 和 z_a ，对两个隐状态进行概率估计 $p(Y|Z_q, Z_a)$
- 图(c)为本文提出的交叉VAE模型，试图通过Q的隐状态 z_q 重构答案A，通过A的隐状态 z_a 重构问题Q，让 z_q 与 z_a 中保留对齐语义，通过这种方式来进行概率估计 $p(Y|Z_q, Z_a)$

How to train?

第一部分：VAE重构损失

式(1)表示用答案a进入Encoder(E_a)得到隐状态 z_a ，并用 z_a 生成问题q

式(2)表示用问题q进入Encoder(E_q)得到隐状态 z_q ，并用 z_q 生成答案a

$$p(q|a) = \mathbb{E}_{z_a \sim p(z_a|a)} [p(q|z_a)], \quad (1)$$

$$p(a|q) = \mathbb{E}_{z_q \sim p(z_q|q)} [p(a|z_q)]. \quad (2)$$

记两个解码器分别为 D_q 和 D_a ，得到两个VAE的重构损失为：

$$\begin{aligned}\mathcal{L}_{cross}(\theta_E, \theta_D) \\ = y \cdot \mathbb{E}_{q \sim Q}[-\log p_D(q|a, E(a))] \\ + y \cdot \mathbb{E}_{a \sim A}[-\log p_D(a|q, E(q))].\end{aligned}\quad (4)$$

- **第二部分：KL散度作为正则化**

变分自编码器采用KL散度正则化器来对齐后验概率 $PE(z_q|q)$ 和概率 $PE(z_a|a)$ ：

$$\begin{aligned}\mathcal{L}_{KL}(\theta_E) = y \cdot \mathbb{E}_{q \sim Q}[D_{KL}(p_E(z_q|q)||p(z_q))] \\ + y \cdot \mathbb{E}_{a \sim A}[D_{KL}(p_E(z_a|a)||p(z_a))],\end{aligned}\quad (5)$$

其中，PE代表后验概率， θ_E 和 θ_D 是需要优化的参数

- **第三部分：匹配结果损失**

匹配结果损失即交叉熵：

$$\begin{aligned}\mathcal{L}_{matching}(\phi_f) = -[y \cdot \log p_{f_\phi}(y|z_q, z_a) \\ + (1 - y) \cdot \log(1 - p_{f_\phi}(y|z_q, z_a))],\end{aligned}\quad (6)$$

- **损失结果：**

$$\mathcal{J} = -\alpha \cdot \mathcal{L}_{cross} - \beta \cdot \mathcal{L}_{KL} + \gamma \cdot \mathcal{L}_{matching}, \quad (7)$$

其中， α, β, γ 为超参数