



基于句法模板采样的无监督复述生成方法

鲍宇¹, 黄书剑^{1,3*}, 周浩², 李磊², 戴新宇¹, 陈家骏¹

1. 南京大学软件新技术国家重点实验室, 南京 210023

2. 字节跳动人工智能实验室, 北京 100086

3. 鹏城实验室, 深圳 518000

* 通信作者. E-mail: huangsj@nju.edu.cn

收稿日期: 2021-02-23; 修回日期: 2021-05-12; 接受日期: 2021-10-25; 网络出版日期: 2022-10-09

国家自然科学基金 (批准号: 6217020152) 和国家重点研发计划前沿科技创新专项 (批准号: 2019QY1806) 资助

摘要 文本复述可以辅助机器翻译、智能问答、文本分类等任务, 是非常重要的自然语言处理任务. 近年来, 一些研究探索了基于结构变换的文本复述, 从无监督学习的概率化表示空间中采样多个句法表示并生成多个复述. 然而, 通过后验分布采样句法表示生成的复述往往高度相似, 缺乏多样性; 另一方面, 从先验分布采样句法表示又难以保证与给定的语义表示相匹配, 导致生成的复述质量欠佳. 本文提出了基于句法模板的文本复述模型, 引入了句法模板隐变量建立语义空间和句法空间的联系, 并进一步提出了两步采样策略: (1) 使用先验分布采样句法模板, 使得采样的句法表示更加多样化; (2) 使用后验分布采样句法表示, 以确保句法表示与语义表示的匹配. 实验表明, 两步采样策略有效地结合了先验采样和后验采样的优势, 生成的文本复述可以在具备良好生成质量的同时保持着更好的多样性, 取得了当前最佳的复述性能.

关键词 无监督复述, 变分自编码器, 句法结构, 采样

1 引言

文本复述 (paraphrasing), 即生成与输入文本语义相同但形式不同的表述, 被广泛地应用于文本分类^[1]、信息检索^[2]、智能问答^[3,4]等下游任务. 早期, 文本复述方案依赖于专家设计的规则或者手工特征^[5,6], 代价昂贵且场景局限. 随着深度学习的发展, 研究人员探索了基于神经网络的复述模型^[7,8], 例如, 利用标注的“输入 – 复述”平行句对, 研究人员可以训练序列到序列 (sequence-to-sequence) 模型完成对输入文本的复述. 然而, 这一类基于有监督学习的方案动辄需要上百万的平行标注数据, 而实际上复述标注数据较为有限, 难以满足这一条件. 此外, 文本复述的一个重要目标是多样性, 而当下有限且单一的平行标注句对难以辅助基于监督学习的复述模型完成这一目标.

引用格式: 鲍宇, 黄书剑, 周浩, 等. 基于句法模板采样的无监督复述生成方法. 中国科学: 信息科学, 2022, 52: 1808–1821, doi: 10.1360/SSI-2021-0065
Bao Y, Huang S J, Zhou H, et al. Unsupervised paraphrasing via syntactic template sampling (in Chinese). Sci Sin Inform, 2022, 52: 1808–1821, doi: 10.1360/SSI-2021-0065

基于无监督学习策略的复述方案有效地缓解了对于标注数据的需求^[9~12]。基于无监督学习策略,研究人员利用大规模的无标注文本建模输入句子的词或者句子的概率分布;随后,通过在建模的概率化空间中采样变换句子^[10]或句中的部分词^[9,11]得到输入文本的复述。除了在词层面进行采样外,研究人员进一步探索了采样句法的复述方案。Bao等^[12]解耦建模了句子的语义和句法的概率化空间(disentangled syntactic and semantic spaces based variational autoencoder, DSS-VAE),并通过采样句法表示变换输入文本的句法生成复述,获得了最佳的复述生成质量。然而,在该方案建模的句法后验分布采样的多个表示之间往往高度相似,尽管语义被有效地保持,但复述的结果缺乏多样性。因此,如何生成高质量且多样化的复述,仍是一个亟待解决的问题。

本文提出了基于句法模板的变分自编码器模型(syntax-template based variational autoencoder, STVAE),通过句法模板控制采样句法表示的过程,生成高质量和多样性的复述。直觉而言,具有相同语义的自然语言常常可以有不同的句型或句式,例如,英文的5种基本句型之间常常可以相互转换,即用于表达相同的语义。相应地,具体的句法表达则应根据语义在不同的句型或句式下使用不同的结构。基于此,STVAE在DSS-VAE模型基础上引入离散的句法模板隐变量建模抽象的句型或句式信息,并假设句法表示的建模依赖于句法模板与语义表示。基于STVAE,本文提出两步采样策略通过句法模板采样多样化的句法表示:首先,从句法模板隐变量的先验分布中采样句法模板;随后,基于语义表示和句法模板,从句法隐变量的后验分布中采样得到句法表示。最后,将句法表示与语义表示结合,生成复述文本。

实验表明,STVAE模型生成的复述文本兼具高质量与多样性,在自动评测以及人工评测指标上均优于此前的基于无监督学习策略的文本复述方案。进一步的分析实验表明,本文引入的句法模板隐变量有效地建模出了句型或句式信息,所提出的两步采样策略有效地利用句法模板结合了先验采样和后验采样的特点,对类似场景具备重要的借鉴意义。

2 相关工作

对于复述生成的探索由来已久。最早,研究人员根据语言学知识设计规则、文法或者浅层的词汇特征构建复述模型^[5,6,13,14],这种方式代价昂贵且场景有限。深度学习时代,研究人员利用大规模标注数据训练基于神经网络的文本生成模型进行复述生成^[7,15,16]。然而,获取大量标注数据的代价是非常昂贵的。因此,研究人员开始探索基于无监督学习策略的复述方案。

基于无监督学习策略的复述方案建模了句子或者词汇的概率分布,并通过设计采样策略变换输入文本的句子或者词生成复述。Miao等^[11]提出从预训练语言模型中使用Metropolis-Hastings采样词汇完成复述。变分自编码器模型^[7,10]可以将句子映射到概率化的隐空间,进而通过直接从输入句子的后验分布采样句子表示生成文本的复述。Bao等^[12]提出对输入文本的句法进行采样变换生成复述,获取了当前最佳生成质量的复述。然而,以上方案均选择从所建模的后验分布进行采样,获取的表示往往与输入文本高度相似,难以达到复述生成的多样化目标。本文提出引入离散的句法模板隐变量,进一步设计了两步采样策略将先验采样的多样性优势与后验采样的高质量优势结合在一起,实现了兼具高质量和多样化的文本复述。

通过提供显式的句法结构^[17]或者隐式的句法参照文本^[18,19],句法转换生成模型可以使用给定的句法信息对语义进行表达,其结果也可以看作是一种基于句法变换的文本复述。然而,为了保证提供的句法信息与语义之间的匹配性,这类方法往往需要精巧的设计,并不适于大规模的应用。与之不同,本文提出的方法可以高效地采样句法表示,并且依然保持着多样性和高质量。

3 背景研究

本文的工作基于变分自编码器模型 (3.1 小节) 以及解耦句法与语义的变分自编码器模型 (3.2 小节), 本节对此进行介绍.

3.1 变分自编码器模型

变分自编码器模型^[10] (variational autoencoder, VAE, 如图 1(a) 所示) 使用一个连续型隐变量 z 建模了句子的隐空间, 将其概率 $p(\mathbf{x})$ 计算为

$$p(\mathbf{x}) = \int p(\mathbf{z})p(\mathbf{x}|\mathbf{z})d\mathbf{z}. \quad (1)$$

使用变分推断技术训练 VAE 模型, 最大化 $p(\mathbf{x})$ 的证据下界 (evidence lower bound, ELBO) 进行参数更新:

$$\begin{aligned} \log p(\mathbf{x}) &\geq \text{ELBO} \\ &= \mathbb{E}_{q(\mathbf{z}|\mathbf{x})} [\log p(\mathbf{x}|\mathbf{z})] - \mathbb{KL}(q(\mathbf{z}|\mathbf{x})\|p(\mathbf{z})), \end{aligned} \quad (2)$$

其中, $p(\mathbf{z})$ 为隐变量的先验分布, $q(\mathbf{z}|\mathbf{x})$ 为隐变量的后验分布, $p(\mathbf{x}|\mathbf{z})$ 则用于建模从隐空间重构出句子 \mathbf{x} 的过程.

3.2 解耦句法和语义的变分自编码器模型

Bao 等^[12] 提出解耦句法和语义信息的变分自编码器模型 (DSS-VAE, 如图 1(b) 所示). 具体地, DSS-VAE 通过假设句法隐变量 \mathbf{s} 与语义隐变量 \mathbf{z} 相互独立, 将 $p(\mathbf{x})$ 建模为

$$p(\mathbf{x}) = \int p(\mathbf{z})p(\mathbf{s})p(\mathbf{x}|\mathbf{z}, \mathbf{s})d\mathbf{z}d\mathbf{s}. \quad (3)$$

相应地, 其证据下界为

$$\begin{aligned} \log p(\mathbf{x}) &\geq \text{ELBO} \\ &= \mathbb{E}_{q(\mathbf{s}|\mathbf{x}), q(\mathbf{z}|\mathbf{x})} [\log p(\mathbf{x}|\mathbf{s}, \mathbf{z})] - \mathbb{KL}(q(\mathbf{s}|\mathbf{x})\|p(\mathbf{s})) - \mathbb{KL}(q(\mathbf{z}|\mathbf{x})\|p(\mathbf{z})), \end{aligned} \quad (4)$$

其中, $p(\mathbf{z})$ 与 $q(\mathbf{z}|\mathbf{x})$ 分别为语义隐变量的先验与后验分布, $p(\mathbf{s})$ 与 $q(\mathbf{s}|\mathbf{x})$ 分别为句法隐变量的先验与后验分布, $p(\mathbf{x}|\mathbf{z}, \mathbf{s})$ 建模了从 \mathbf{z} 以及 \mathbf{s} 重构出输入句子 \mathbf{x} 的过程.

在生成复述时, DSS-VAE 模型通过后验分布 $q(\mathbf{s}|\mathbf{x})$ 采样句法表示后, 与语义表示组合得到文本的复述. 该方案可以极大地保留语义信息, 实现了最佳的生成质量. 然而, 其句法的后验分布 $q(\mathbf{s}|\mathbf{x})$ 依赖于给定的输入 \mathbf{x} , 采样出的多个句法表示往往高度相似, 生成的复述多样性较差. 此外, DSS-VAE 模型在句法迁移生成任务中的实践表明: 句法表示与语义表示需要满足一定的匹配性, 否则很难生成正确的句子^[12]. 例如, 不同长度的句子提取的句法表示和语义表示组合生成的句子从语义上和句法上均不能满足要求.

4 基于句法模板采样的无监督复述模型

人类说话的方式需遵循一定的句法, 这种句法以一种形式的且不受语境影响并带有转换生成规则的语法^[20] 为特征. 基于此, DSS-VAE 模型提出分离建模自然语言的句法和语义信息, 通过保留语义

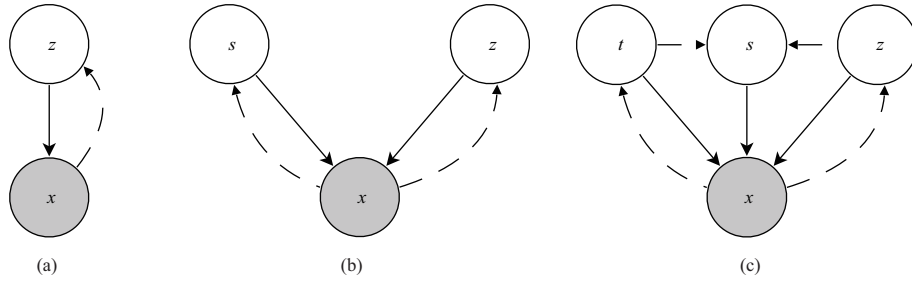


图 1 (a) VAE, (b) DSS-VAE 和 (c) STVAE 的图模型示意图
 Figure 1 Graphical models of (a) VAE, (b) DSS-VAE, and (c) STVAE

而变换句法生成高质量的复述. 然而, 使用输入文本 x 作为条件, 难以脱离输入文本的句法从近似后验分布 $q(s|x)$ 采样出多样化的句法表示. 针对于此, 本文提出基于句法模板的变分自编码器模型 (STVAE), 并探索了一种更为灵活的采样策略用于生成多样化的文本复述.

4.1 基于句法模板的句法与语义解耦建模

尽管自然语言所包含的句法结构难以穷尽枚举, 但语言学家们归类的句型、句式等抽象的句法结构类型是有限的. 例如, 英语的基本句型仅有 5 种, 并且不同句型之间存在转换的可能, 同样一个语义使用不同句型会有不同形式的句法表达. 本文提出引入离散的句法模板隐变量对抽象的句法信息 (例如: 句型、句式) 进行建模, 并相应地让具体的句法表达形式由语义与句法模板决定.

如图 1(c) 所示, 句法隐变量 s 依赖于句法模板隐变量 t 和语义隐变量 z . 因此, STVAE 一方面可以通过变换句法模板来变换句法表示; 另一方面更建立了语义表示与句法表示的关联, 保障了句法表示与语义表示的匹配. STVAE 模型首先将句子的概率 $p(x)$ 建模为

$$p(x) = \int p(t, z, s)p(x|t, z, s)dt dz ds. \quad (5)$$

随后假设了语义 z 和 t 之间的独立 (支持句法变换), 引入 $p(x|t, z, s)$ 对应的近似后验分布:

$$q_\phi(t, z, s|x) = q_\phi(t|x)q_\phi(z|x)q_\phi(s|t, z). \quad (6)$$

相应地, STVAE 模型的证据下界为

$$\begin{aligned} \log p_\theta(x) &\geq \text{ELBO} \\ &= \mathbb{E}_{q_\phi(t, z, s|x)} [\log p_\theta(x|t, z, s)] \\ &\quad - \mathbb{KL}(q_\phi(t|x)||p_\theta(t)) - \mathbb{KL}(q_\phi(z|x)||p_\theta(z)) - \mathbb{KL}(q_\phi(s|z, t)||p_\theta(s)), \end{aligned} \quad (7)$$

其中, 先验分布 $p_\theta(s)$ 和 $p_\theta(z)$ 是参数为 $\mathcal{N}(0, \mathbf{I})$ 的标准高斯分布 (Gaussian distribution), 相应的近似后验分布 $q_\phi(s|x)$ 和 $q_\phi(z|x)$ 是参数分别为 $\mathcal{N}(\mu_s, \sigma_s^2)$ 与 $\mathcal{N}(\mu_z, \sigma_z^2)$ 的高斯分布; t 为包含 T 个类别中的离散隐变量, 其先验分布 $p_\theta(t) = 1/T$, 其后验分布 $q_\phi(t|x)$ 则为参数 ϕ 定义的种类分布 (categorical distribution). 参数化的过程通过推断网络完成.

4.2 STVAE 模型参数化

本文使用神经网络模型对 STVAE 模型中的各个隐变量的近似后验分布进行参数化. 首先, 使用循环神经网络 (recurrent neural networks, RNN)^[21] 定义的编码器获取句子的表示 r_x ; 随后, 将其划分

成两部分 $\mathbf{r}_x = [\mathbf{r}_x^{\text{sem}}; \mathbf{r}_x^{\text{syn}}]$, 分别用于对语义信息和句法信息进行建模.

语义后验分布推断. 给定编码的语义信息 $\mathbf{r}_x^{\text{sem}}$, 利用多层感知机 (multi-layer perceptron, MLP) 网络计算语义隐变量 z 的近似后验分布 $q_\phi(z|\mathbf{x})$ 的参数 μ_z 与 σ_z :

$$\begin{bmatrix} \mu_z \\ \sigma_z \end{bmatrix} = \begin{bmatrix} W_z^\mu \\ W_z^\sigma \end{bmatrix} \text{ReLU}(W_z \mathbf{r}_x^{\text{sem}} + \mathbf{b}_z), \quad (8)$$

其中, ReLU (rectified linear unit)^[22] 为激活函数, W_z^μ , W_z^σ , W_z 和 \mathbf{b}_z 为模型参数.

句法后验分布推断. 给定编码的句法信息 $\mathbf{r}_x^{\text{syn}}$, 对句法模板隐变量以及句法隐变量的近似后验分布进行推断. 首先, 将离散的句法模板隐变量 t 的近似后验分布 $q_\phi(t|\mathbf{x})$ 的计算为

$$q_\phi(t|\mathbf{x}) = \text{Softmax}(W_2^t \text{ReLU}(W_1^t \mathbf{r}_x^{\text{syn}} + \mathbf{b}_1^t) + \mathbf{b}_2^t), \quad (9)$$

其中, W_1^t , W_2^t , \mathbf{b}_1^t 和 \mathbf{b}_2^t 是模型的参数. 利用最大化后验分布采样得到的语义隐变量 $z^* = \mu_z$ 以及句法模板隐变量 $t^* = \arg \max_t q_\phi(t|\mathbf{x})$, 计算句法隐变量 s 的近似后验分布 $q_\phi(s|z; t)$ 的参数 μ_s 与 σ_s :

$$\begin{bmatrix} \mu_s \\ \sigma_s \end{bmatrix} = \begin{bmatrix} W_s^\mu \\ W_s^\sigma \end{bmatrix} \text{ReLU}(W_s [z^*; \text{one-hot}(t^*)] + \mathbf{b}_s), \quad (10)$$

其中, W_s^μ , W_s^σ , W_s 和 \mathbf{b}_s 是模型的参数.

解码器. 从隐变量重构出输入文本的模型 $p_\theta(\mathbf{x}|t, z, s)$ 通常由循环神经网络进行参数化, 将隐变量 t, z , 以及 s 拼接作为初始状态, 逐词解码出 \mathbf{x} :

$$p_\theta(\mathbf{x}|t, z, s) = \prod_t^{|\mathbf{x}|} p_\theta(\mathbf{x}_t | \mathbf{x}_{<t}; t, z, s), \quad (11)$$

$$p_\theta(\mathbf{x}_t | \mathbf{x}_{<t}; t, z, s) = \text{RNN}([\mathbf{x}_{<t}; z; s; \mathbf{E}_t]),$$

其中 \mathbf{E} 为离散的句法模板隐变量的嵌入表示.

4.3 STVAE 模型的学习

STVAE 模型通过多任务学习以及对抗学习策略设置了针对句法信息以及语义信息的损失项, 从而完成对句法和语义的解耦学习.

句法和语义隐空间的解耦建模. 如图 2 所示, 参考 Bao 等^[12], 本文使用循环神经网络预测序列化的成分句法树^[23] ($\mathbf{o} = [o_1, o_2, \dots, o_n]$) 以建模句法信息, 使用多层感知机预测句子的词袋分布 (bag-of-words distribution, $\mathbf{y} = \{y_1, y_2, \dots, y_m\}$) 以建模语义信息.

除优化证据下界外, STVAE 模型利用多任务学习 (multi-task learning) 对句法信息和语义信息进行建模, 训练句法表示建模句法信息, 训练语义表示建模语义信息:

$$\mathcal{L}_{\text{sem}}^{(\text{mul})} = - \sum_{w \in \mathcal{V}} y_w \log p(w|z), \quad (12)$$

$$\mathcal{L}_{\text{syn}}^{(\text{mul})} = - \sum_{i=1}^{|\mathbf{o}|} \log p(o_i | o_1 \cdots o_{i-1}, \mathbf{s}, t).$$

随后, 使用对抗训练 (adversarial training) 技术分离出句法表示中的语义信息以及语义表示中的句法信息. 具体地, 训练一个语义对抗器从句法表示中预测出词袋分布, 训练一个句法对抗器从语义表

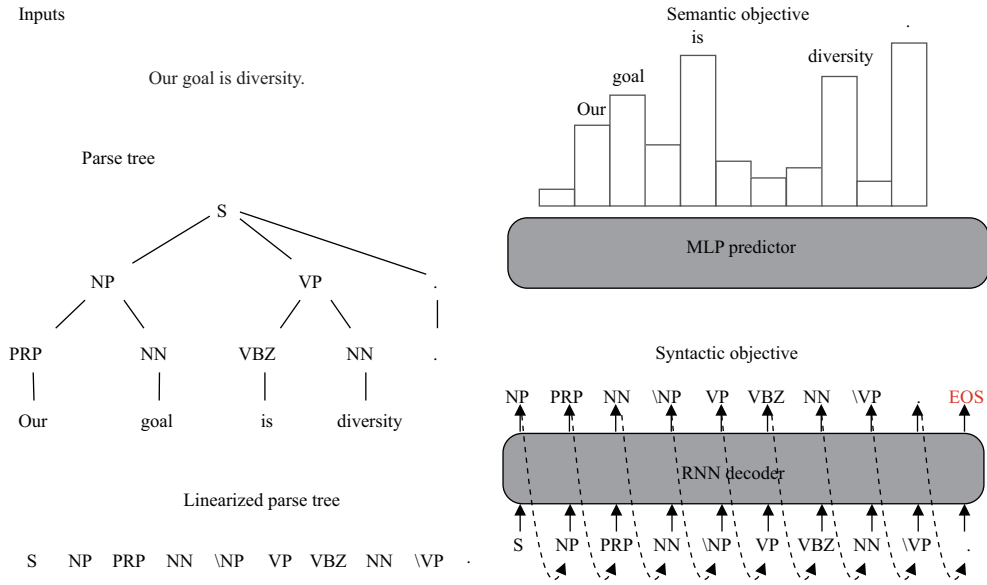


图 2 STVAE 模型的语义与句法信息的建模目标

Figure 2 The syntactic and semantic objective of STVAE

示预测句法树序列. 在优化 STVAE 时, 计算以上对抗器的损失, 并作为负增益项来对抗句法表示中的语义信息以及语义表示中的句法信息.

$$\begin{aligned} \mathcal{L}_{\text{sem}}^{(\text{adv})} &= \sum_{w \in \mathcal{V}} y_w \log p_{\text{adv}}(w | \mathbf{s}, \mathbf{t}), \\ \mathcal{L}_{\text{syn}}^{(\text{adv})} &= \sum_{i=1}^{|\mathbf{o}|} \log p_{\text{adv}}(o_i | o_1 \cdots o_{i-1}, \mathbf{z}). \end{aligned} \tag{13}$$

此外, 为了确保句法和语义表示都能够在生成文本时发挥作用, 还需训练对抗器引入对抗损失来阻止句法和语义表示单独重构出文本:

$$\begin{aligned} \mathcal{L}_{\text{rec}}^{(\text{adv})}(\mathbf{z}) &= \sum_{i=1}^{|\mathbf{x}|} \log p_{\text{rec}}(x_i | x_{<i}, \mathbf{z}), \\ \mathcal{L}_{\text{rec}}^{(\text{adv})}(\mathbf{s}, \mathbf{t}) &= \sum_{i=1}^{|\mathbf{x}|} \log p_{\text{rec}}(x_i | x_{<i}, \mathbf{s}, \mathbf{t}). \end{aligned} \tag{14}$$

需要说明的是, 对抗器训练时的梯度不会回传给 STVAE 进行参数更新, 其训练过程与 STVAE 模型的训练是交替且独立的.

句法模板的学习. 离散的句法模板隐变量用于建模抽象的句法信息 (句型或者句式等), 为了使学出的句法模板更加多样, 本文引入一个正则化项损失确保模板表示的多样性:

$$\mathcal{L}_{\text{div}} = \sum_{1 \leq i < j \leq T} \text{dist}(\mathbf{E}_i, \mathbf{E}_j), \tag{15}$$

其中, \mathbf{E}_i 以及 \mathbf{E}_j 分别是第 i 和 j 个的模板表示, $\text{dist}(\cdot, \cdot)$ 计算了两者之间的距离. 本文选用余弦距离度量表示之间的距离 (距离越大, 两者越相似), 并通过最小化该损失项来学习多样化的表示.

全局目标函数. 综合前文所提出的解耦建模句法和语义的目标函数、学习句法模板的目标函数, 以及 STVAE 模型的证据下界, 得到全局目标函数:

$$\begin{aligned}
 \mathcal{L} = & - \mathbb{E}_{q(\mathbf{z}, \mathbf{t}, \mathbf{s} | \mathbf{x})} \log[p(\mathbf{x} | \mathbf{z}, \mathbf{s}, \mathbf{t})] \\
 & + \lambda_{\text{sem}}^{\text{KL}} \mathbb{KL}(q(\mathbf{z} | \mathbf{x}) \| p(\mathbf{z})) + \lambda_{\text{syn}}^{\text{KL}} \mathbb{KL}(q(\mathbf{t} | \mathbf{x}) \| p(\mathbf{t})) + \lambda_{\text{syn}}^{\text{KL}} \mathbb{KL}(q(\mathbf{s} | \mathbf{z}, \mathbf{t}) \| p(\mathbf{s})) \\
 & + \lambda_{\text{sem}}^{\text{mul}} \mathcal{L}_{\text{sem}}^{\text{(mul)}} + \lambda_{\text{sem}}^{\text{adv}} \mathcal{L}_{\text{sem}}^{\text{(adv)}} + \lambda_{\text{sem}}^{\text{rec}} \mathcal{L}_{\text{sem}}^{\text{(adv)}}(\mathbf{z}) + \lambda_{\text{syn}}^{\text{mul}} \mathcal{L}_{\text{syn}}^{\text{(mul)}} + \lambda_{\text{syn}}^{\text{adv}} \mathcal{L}_{\text{syn}}^{\text{(adv)}} + \lambda_{\text{syn}}^{\text{rec}} \mathcal{L}_{\text{syn}}^{\text{(adv)}}(\mathbf{s}, \mathbf{t}) \\
 & + \lambda_{\text{div}} \mathcal{L}_{\text{div}}, \tag{16}
 \end{aligned}$$

其中, $\lambda_{\text{sem}}^{\text{KL}}$, $\lambda_{\text{syn}}^{\text{KL}}$, $\lambda_{\text{sem}}^{\text{mul}}$, $\lambda_{\text{sem}}^{\text{adv}}$, $\lambda_{\text{sem}}^{\text{rec}}$, $\lambda_{\text{syn}}^{\text{mul}}$, $\lambda_{\text{syn}}^{\text{adv}}$, $\lambda_{\text{syn}}^{\text{rec}}$ 和 λ_{div} 为超参数, 用于调整各项损失在全局目标函数中的权重.

4.4 基于 STVAE 模型的两步采样策略

在生成复述时, STVAE 模型通过句法模板控制句法表示的采样过程, 基于输入文本语义表示, 使用不同的句法模板得到不同的句法表示, 进而生成多样化的复述.

- **先验分布采样.** 从句法模板的先验分布中采样出 N 个不同的句法模板, 保证句法表示的多样性: $\mathbf{t} \sim \text{Categorical}[0, T]$.
- **最大化后验分布采样.** 使用语义表示和句法表示的近似后验分布采样, 从而确保句法表示与语义表示的匹配性:

$$\begin{aligned}
 \mathbf{z} &= \boldsymbol{\mu}_{\mathbf{z}}, \\
 \mathbf{s} &= W_s^{\mu} \text{ReLU}(W_s[\boldsymbol{\mu}_{\mathbf{z}}; \text{one-hot}(\mathbf{t})] + \mathbf{b}_s),
 \end{aligned}$$

其中 \mathbf{t} 为采样的句法模板.

基于语义表示 \mathbf{z} 、句法表示 \mathbf{s} , 以及句法模板 \mathbf{t} , 使用解码器生成文本的复述.

5 实验

5.1 实验设置

数据集设置. 为了分析 STVAE 的复述性能, 实验选取了与文献 [24] 相同的数据集 (Quora¹) 和 Twitter^[25]) 及数据划分. 其中, Quora 数据集提供了约 140000 条标注的复述句对, 其中 100000 条数据作为训练集合, 3000 条句对作为验证集合用于挑选模型, 30000 条句对用于测试模型的性能; Twitter 数据集为标注了互为复述的 twitter, 包含 5000 和 1000 条由人工标注的测试集和验证集, 以及 110000 条自动标注的平行句对作为训练集. 在训练 STVAE 时, 所有的平行句对均拆分成非平行句以进行无监督学习. 以上数据的句法树标注则通过句法分析器 ZPar²) 完成.

基线系统. 为了验证 STVAE 模型的有效性, 实验选取 VAE 模型^[10] 以及 DSS-VAE 模型^[12] 等基于无监督学习的模型作为基线系统. 其中, DSS-VAE 模型为当前复述质量最佳的无监督复述方法.

此外, 实验还选取了一系列基于有监督模型作为参照: 基于柱搜索的序列到序列模型 (Seq2seq-BS)、条件变分自编码器 (VAE-SVG-eq)^[7]、基于多解码器词嵌入的生成器 (D-PAGE)^[26]、基于指针

1) <https://www.kaggle.com/c/quora-question-pairs/data>.

2) <https://www.sutd.edu.sg/cmsresource/faculty/yuezhang/zpar.html>.

网络的复述模型 (PG-BS) [27] 和基于强化学习策略的多生成器模型 (multi-generator) [24]. 其中, multi-generator 模型是当前获取最佳多样性的有监督复述模型.

超参数设置及实现. Quora 和 Twitter 数据集任务中的词表分别设置为高频的 30000 和 40000 个词. 隐变量表示的维度设置为 128, 词嵌入表示的维度则设置为 256. 参考文献 [12], 将训练 STVAE 的优化目标公式 (16) 中的超参数设为 $\lambda_{\text{sem}}^{\text{KL}} = 0.67$, $\lambda_{\text{syn}}^{\text{KL}} = 0.33$, $\lambda_{\text{sem}}^{\text{mul}} = \lambda_{\text{sem}}^{\text{adv}} = 1.0$, $\lambda_{\text{sem}}^{\text{rec}} = \lambda_{\text{syn}}^{\text{mul}} = \lambda_{\text{syn}}^{\text{adv}} = \lambda_{\text{syn}}^{\text{rec}} = 0.1$, 模板多样性的损失权重 λ_{div} 则设为 0.1³⁾.

实验使用 Adam [28] ($\beta_1 = 0.9$, $\beta_2 = 0.995$) 优化模型, 并将初始学习率设置为 0.001. 参考文献 [10], 为了避免 VAE 模型出现模式坍塌 (mode collapse) 问题, 模型的训练使用了 KL 权重退火以及随机词替换策略. KL 项的权重按照 sigmoid 函数从 0 逐步增加至预先设定的 $\lambda_{\text{sem}}^{\text{KL}}$ 和 $\lambda_{\text{syn}}^{\text{KL}}$; 为了避免隐变量被解码器忽视, 训练时会将解码器的输入按照 0.5 的概率替换为未登录词 <unk>.

评价指标. 复述的评价则从质量和多样性两方面进行. 具体地, 对于测试集中的每一个句子, 通过复述模型生成 N (本文设 $N = 3$) 个对应的复述候选, 并计算如下指标评测复述的质量以及多样性:

- **BLEU 和 METEOR.** BLEU [29] 可以反映输出复述与给定参考复述之间的词匹配程度, 而 METEOR [30] 则引入词干以及同义词信息, 从语义角度进一步对输出的质量进行了评测. 具体地, 计算 N 个复述输出的 BLEU 或 METEOR 的平均得分作为每个样本的得分. 最终的得分为所有样本的平均得分, 得分越高, 表明生成质量越好.

- **self-BLEU.** 参考文献 [24], 实验使用 self-BLEU 评测复述的多样性. self-BLEU 通过计算每个样本的 N 个候选复述之间的相似度, 反映了生成复述的多样性. 最终的得分为所有样本的平均得分, 得分越低, 则表明生成多样性越好.

本文进一步使用生成质量与多样性得分的比值 BPD (BLEU per diversity) 和 MPD (METEOR per diversity) 反映复述的综合性能:

$$\text{BPD} = \frac{\text{BLEU4}}{\text{self-BLEU4}} \times 100\%, \quad \text{MPD} = \frac{\text{METEOR}}{\text{self-BLEU4}} \times 100\%.$$

得分越高, 表明生成的复述单位多样性得分对应的质量得分更高, 即复述的综合性能越好. 需要说明的是, 生成多样性仅当生成质量满足一定水平时才有意义, 因此, BPD 和 MPD 指标更加适用于模型具备高生复述质量的情况 (有关 BPD 以及 MPD 指标的分析见实验).

5.2 实验结果

自动评测. 表 1 是各个模型在 Quora 数据集上的评测结果. 显然, 在所有基于无监督学习策略的复述模型中, STVAE 实现了最佳的综合复述性能. 通过分离句法信息和语义信息, STVAE 和 DSS-VAE 有效地保留了输入文本的语义信息, 实现了与有监督方法相近的复述质量 (METEOR).

进一步对比基于无监督的方法, 可以发现: 生成的复述在质量和多样性存在着一定的权衡. DSS-VAE 模型生成的复述具备着最佳的生成质量, 但多样性却相对较差; 与之相对应地, STVAE 模型则生成了更多样化的复述, 但质量却相对较差. 使用本文提出的 BPD 和 MPD 对这种权衡进行分析: STVAE 模型获取了更高的 BPD 和 MPD 得分, 表明 STVAE 通过引入句法模板隐变量以及两步采样策略有效地平衡了复述的质量与多样性, 是一种综合性能更佳的复述方案.

此外, 可以发现: 与基于有监督的复述方案对比, 无监督复述模型可以实现较为接近的 METEOR 得分, 但 BLEU 得分悬殊较大. 然而, 较低的 BLEU 得分并不意味着复述的生成质量较差. BLEU 得分计算的是与给定的参考复述之间的匹配程度, 而无监督的复述方案通过变换输入文本得到复述文本,

3) 模型的实现和更多实验细节见 <https://github.com/baoy-nlp/STVAE>.

表 1 Quora 数据集上的性能^{a)}
Table 1 Performances on Quora dataset^{a)}

Model	Quality				Diversity			BPD [†]	MPD [†]
	BLEU2 [†]	BLEU3 [†]	BLEU4 [†]	METEOR [†]	self-BLEU2 [‡]	self-BLEU3 [‡]	self-BLEU4 [‡]		
Supervised methods									
Seq2seq-BS [†]	37.03	27.53	21.11	27.78	73.89	66.51	59.33	35.58	46.82
VAE-SVG-eq [†]	34.78	25.96	19.97	26.45	89.82	87.53	85.43	23.38	30.96
PG-BS [†]	38.91	29.26	22.54	28.88	75.60	68.72	61.89	36.42	46.66
Multi-generator [†]	36.84	27.46	20.99	29.28	56.38	47.76	40.55	51.76	72.21
Unsupervised methods									
VAE	16.04	9.92	5.56	27.65	67.90	58.34	48.08	11.56	57.32
DSS-VAE	22.01	14.07	8.82	35.40	92.45	90.39	87.52	10.08	40.45
STVAE	17.46	10.21	6.00	29.78	58.37	46.09	36.36	16.50	81.90

a) The larger[†] (or lower[‡]), the better. [†]: the results are quoted from [24].

并不服务于给定的参考复述. 而通过引入词干替换等扩充单一复述对生成质量进行评测的 METEOR 得分与有监督复述方案接近, 表明基于无监督学习生成的复述质量是有保障的.

表 2 给出了各个模型在 Twitter 数据集上的结果. 相较于 Quora, Twitter 上的实验结果更加凸显出复述生成在质量与多样性之间的权衡. 与此同时, STVAE 模型仍然取得了最高的 BPD 得分和 MPD 得分, 表明该模型在 Twitter 数据集中同样实现了最佳的复述性能.

人工评测. 除自动评测外, 实验进一步通过人工评测了生成的复述. 邀请 3 位英语的熟练使用者⁴⁾从流利度和相关性两个维度评测随机选取的 100 个样本对应的复述, 按流利度 (或相关性) 对各模型的输出进行排序, 流利度越高 (或相关性越高), 则对应复述的得分越高.

表 3 为基于无监督方法的复述模型的人工评测得分, STVAE 模型取得了最佳的复述性能. 该结果表明: 基于句法模板隐变量和两步采样策略, 有效地保障了句法表示的多样性以及与语义表示的匹配度, 进而保证了最终生成复述的综合性能. 此外, DSS-VAE 和 STVAE 模型均生成了比 VAE 模型更流利且相关的复述, 则进一步验证了建模句法以及解耦语义对于生成复述的作用.

5.3 分析与讨论

两步采样策略的有效性分析. 表 4 对比了不同采样策略的变分自编码器模型在 Quora 与 Twitter 上的性能. 显然, 由于先验分布并不包含特定的输入信息, 尽管采样得到的样本极为多样化, self-BLEU 得分极低, 但极低的 METEOR 得分则表明这种采样本身无法完成复述目标. 另一方面, 尽管最大化近似后验分布采样 (max A posterior sampling) 并生成复述文本可以获得较高质量的复述, 但其多样性则同样无法满足复述目标. 基于 STVAE, 本文提出地两步采样策略综合了先验分布采样和后验分布采样的优势: 利用先验分布采样句法模板保障句法表示的多样性, 与此同时, 利用后验分布采样句法表示保障生成复述的质量.

BPD 与 MPD 指标的适用性分析. 结合文献 [24] 以及本文在 Twitter 数据集所进行的人工评测结果, 表 5 给出了 BPD 以及 MPD 与人工评测指标的皮尔逊相关系数 (Pearson correlation coefficient). 可以发现, BPD 相比较 MPD 在人工评测的流利度与相关度上具备着更高的相关性 (皮尔逊相关系数 >0.7); 而 MPD 有监督场景和无监督场景则体现出了一定的差异: 在有监督场景, MPD 与流利度与相

4) 通过了 CET6 (大学生英语六级) 考试.

表 2 Twitter 数据集上的性能 ^{a)}
Table 2 Performances on Twitter dataset^{a)}

Model	Quality				Diversity			BPD [†]	MPD [†]
	BLEU2 [†]	BLEU3 [†]	BLEU4 [†]	METEOR [†]	self-BLEU2 [↓]	self-BLEU3 [↓]	self-BLEU4 [↓]		
Supervised methods									
Seq2seq-BS [†]	32.69	28.25	24.99	22.51	82.79	80.29	77.98	32.05	28.87
VAE-SVG-eq [†]	26.43	23.04	20.57	18.34	91.46	90.51	89.67	22.94	20.45
D-PAGE [†]	32.95	28.82	25.88	22.59	88.35	86.45	84.76	30.53	26.65
PG-BS [†]	33.86	29.42	26.15	23.52	82.57	79.99	77.48	33.75	30.36
Multi-generator [†]	34.23	29.66	26.38	24.29	65.83	61.17	57.45	45.92	42.28
Unsupervised methods									
VAE	13.88	9.61	7.62	21.32	47.19	43.32	39.92	19.09	53.41
DSS-VAE	21.21	16.11	12.82	33.06	76.63	72.21	68.95	18.59	47.59
STVAE	17.95	13.87	11.42	28.38	55.41	51.51	49.66	23.00	57.15

a) The larger[†] (or lower[↓]), the better. †: the results are quoted from [24].

表 3 模型在 Twitter 数据集的人工评测打分 ^{a)}
Table 3 Human evaluation scores on the Twitter dataset^{a)}

Model	Fluency [†]	Relevance [†]
VAE	2.17 (±0.03)	2.12 (±0.05)
DSS-VAE	2.31 (±0.01)	2.34 (±0.05)
STVAE	2.43 (±0.01)	2.69 (±0.03)

a) The larger[†], the better.

表 4 不同采样方法的性能 ^{a)}
Table 4 Performance of different sampling methods^{a)}

Methods	Quora		Twitter		
	METEOR [†]	self-BLEU4 [↓]	METEOR [†]	self-BLEU4 [↓]	
VAE	Prior	5.72	0.18	3.56	0.39
	Posterior	27.65	48.08	21.32	39.92
DSS-VAE	Prior	18.77	6.37	12.25	2.78
	Posterior	35.40	87.52	33.06	68.95
STVAE	Two-step	29.78	36.36	28.38	49.66

a) The larger[†] (or lower[↓]), the better.

表 5 Twitter 数据集上 BPD (或 MPD) 与人工评估的相关性分析
Table 5 Correlation analysis of BPD (or MPD) to human evaluation on Twitter

Metrics	Supervised setting		Unsupervised setting	
	Fluency	Relevance	Fluency	Relevance
BPD	0.8005	0.9453	0.8005	0.9453
MPD	0.8132	0.9492	0.8132	0.9492

表 6 STVAE 模型在 Quora 数据集学习的句法模板样例
Table 6 Examples of the syntactic template on Quora dataset

Template number	Examples
1	If the sun is in space, why is the universe so dark?
	If you could go back in time and redo one thing, what would it be?
	If you were the president of the United States, what would you do?
2	Who are some leading Indian female writers?
	Who are the most irritating actors and actresses in bollywood?
	What are the best books for clat 2017?
3	I am 24. Is it too late to get into medicine?
	I want to learn digital marketing. Where do I start?
	Entrepreneurship in India: any new business ideas with lesser investment?
4	How do I raise funds for startup?
	How do I log out from quora website?
	How do you value a business?
5	Should I take a new job or go for solo travelling?
	Should I sell an iphone 6s and buy an iphone se?
	Can you get a second hand high from smoke outside?

关度上同样具备高相关性, 而在无监督场景, 则为低相关性 (皮尔逊系数 <0.5) 或者中相关度 (皮尔逊系数 <0.7). 综合而言, 当生成质量得分较高时, BPD 与 MPD 与人工评测结果是高相关的, 即适合于自动评估复述的综合性能, BPD 与 MPD 得分越高的复述模型综合性能越佳.

句法模板样例分析. 表 6 是从 Quora 数据中所学习出的句法模板的一些样例, 可以看出 STVAE 有效的建模出了诸如句型等抽象的句法结构信息: 1 号模板均为包含 If 条件状语的疑问句; 编号为 2 的模板对应着由 Who 或 What 引导的疑问句; 3 号模板对应复合句类型; 编号为 4 和 5 的句法模板则分别对应着由 How 引导的疑问句以及由情态动词引导的疑问句.

生成的复述样例分析. 表 7 为 DSS-VAE 和 STVAE 模型在 Quora 数据集的开发集上生成的复述样例, 可以发现 STVAE 模型相对于 DSS-VAE 模型生成了兼具多样性与质量的复述文本:

- DSS-VAE 模型的输出不仅相互之间高度相似, 同时与输入文本也十分相似, 而 STVAE 模型生成的复述则更加多样化, 包含了更多的位置替换和句法转换.
- 通过解耦出语义信息作为输入, DSS-VAE 和 STVAE 模型均按照预期地保存了输入文本的语义信息 – 即词袋信息.

此外, 值得注意的是, 目前的变换较为缺乏词的变换, 这可能是由于词袋作为语义建模的目标过于严苛, 潜在地限制了变换句法结构的空間.

6 结论

文本复述的目标是生成多样化且与输入语义相同的文本. 为了这一目标, 本文提出 STVAE 模型, 引入句法模板隐变量建模抽象的句法结构信息, 并设计了两步采样策略获取句法表示生成文本的复述: 利用先验分布采样的多样化特点对句法模板进行采样, 保障了复述的多样性; 基于语义和句法模板, 使用建模的后验分布对句法表示进行采样, 保障了复述的生成质量. 实验表明, STVAE 模型在保持良好

表 7 DSS-VAE 与 STVAE 的复述样例
Table 7 Paraphrasing examples of DSS-VAE and STVAE models

Input#1	Which books should entrepreneurs read?
Output#1	DSS-VAE
	Which books should entrepreneurs read?
	Which books should entrepreneurs read?
Output#1	STVAE
	Which books should entrepreneurs read?
	What books should entrepreneurs read?
	Which books are the entrepreneurs to read?
Input#2	What can I do to get better at basketball?
Output#2	DSS-VAE
	What can I do to get into 10 minutes into basketball?
	What can I do to get more into 10, basketball?
Output#2	STVAE
	What can I do to get better at basketball?
	How can I prepare to get better at basketball?
	What should I do to make basketball better?
Input#3	Is vacuum energy infinite? If it is, how and why? Is it dark energy?
Output#3	DSS-VAE
	Is vacuum energy infinite? If it is, how is it? And why is it?
	Is vacuum energy infinite? If it is, how is it? How and why not?
Output#3	STVAE
	Is vacuum energy infinite? If it is, how is it? And why is it?
	Is vacuum energy infinite? If it is, how and why? Is it dark energy?
	Why is vacuum energy infinite? If it is, and how is it dark energy?
	How is vacuum is and is it infinite? How is it dark energy? How?

的复述质量地同时显著提升了复述的多样性, 获取了最佳的复述综合性能。

样例分析表明, 使用词袋作为语义建模目标过于严格, 一定程度上限制了句法变换的空间。在未来工作中, 将探索更适用于句法变换的语义建模目标, 使得模型可以完成更加高质量且多样化的复述。

参考文献

- 1 Petrović S, Osborne M, Lavrenko V. Using paraphrases for improving first story detection in news and Twitter. In: Proceedings of Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2012. 338–346
- 2 Knight K, Marcu D. Statistics-based summarization-step one: sentence compression. In: Proceedings of the 17th National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence, 2000. 703–710
- 3 Fader A, Zettlemoyer L, Etzioni O. Open question answering over curated and extracted knowledge bases. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2014. 1156–1165
- 4 Yin P C, Duan N, Kao B, et al. Answering questions with complex semantic constraints on open knowledge bases. In: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, 2015. 1301–1310
- 5 McKeown K. Paraphrasing questions using given and new information. *Am J Comput Linguist*, 1983, 9: 1–10
- 6 Ellsworth M, Janin A. Mutaphrase: paraphrasing with framenet. In: Proceedings of ACL-PASCAL Workshop on Textual Entailment and Paraphrasing, 2007. 143–150

- 7 Gupta A, Agarwal A, Singh P, et al. A deep generative framework for paraphrase generation. In: Proceedings of AAAI Conference on Artificial Intelligence, 2018
- 8 Li Z C, Jiang X, Shang L F, et al. Paraphrase generation with deep reinforcement learning. In: Proceedings of Conference on Empirical Methods in Natural Language Processing, 2018. 3865–3878
- 9 Liu X G, Mou L L, Meng F D, et al. Unsupervised paraphrasing by simulated annealing. 2019. ArXiv:1909.03588
- 10 Bowman S, Vilnis L, Vinyals O, et al. Generating sentences from a continuous space. In: Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning, 2016. 10–21
- 11 Miao N, Zhou H, Mou L L, et al. CGMH: constrained sentence generation by metropolis-hastings sampling. In: Proceedings of AAAI Conference on Artificial Intelligence, 2019. 6834–6842
- 12 Bao Y, Zhou H, Huang S J, et al. Generating sentences from disentangled syntactic and semantic spaces. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019. 6008–6019
- 13 Narayan S, Reddy S, Cohen S B. Paraphrase generation from latent-variable PCFGs for semantic parsing. In: Proceedings of the 9th International Natural Language Generation Conference, 2016. 153–162
- 14 Zhao S Q, Lan X, Liu T, et al. Application-driven statistical paraphrase generation. In: Proceedings of Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, 2009. 834–842
- 15 Prakash A, Hasan S A, Lee K, et al. Neural paraphrase generation with stacked residual LSTM networks. In: Proceedings of the 26th International Conference on Computational Linguistics: Technical Papers, 2016. 2923–2934
- 16 Wang S, Gupta R, Chang N, et al. A task in a suit and a tie: paraphrase generation with semantic augmentation. In: Proceedings of AAAI Conference on Artificial Intelligence, 2019, 33. 7176–7183
- 17 Iyyer M, Wieting J, Gimpel K, et al. Adversarial example generation with syntactically controlled paraphrase networks. In: Proceedings of Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2018. 1875–1885
- 18 Chen M D, Tang Q M, Wiseman S, et al. Controllable paraphrase generation with a syntactic exemplar. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019. 5972–5984
- 19 Zhang X Y, Yang Y, Yuan S Y, et al. Syntax-infused variational autoencoder for text generation. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019. 2069–2078
- 20 Chomsky N. Syntactic Structures. Berlin: Walter de Gruyter. 2002
- 21 Cho K, van Merriënboer B, Gulcehre C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In: Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014. 1724–1734
- 22 Nair V, Hinton G E. Rectified linear units improve restricted Boltzmann machines. In: Proceedings of the 27th International Conference on Machine Learning, 2010
- 23 Vinyals O, Kaiser L, Koo T, et al. Grammar as a foreign language. In: Proceedings of the 28th International Conference on Neural Information Processing Systems, 2015. 2773–2781
- 24 Qian L H, Qiu L, Zhang W N, et al. Exploring diverse expressions for paraphrase generation. In: Proceedings of Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 2019. 3164–3173
- 25 Lan W W, Qiu S Y, He H, et al. A continuously growing dataset of sentential paraphrases. In: Proceedings of Conference on Empirical Methods in Natural Language Processing, 2017. 1224–1234
- 26 Xu Q K, Zhang J Y, Qu L Z, et al. D-page: diverse paraphrase generation. 2018. ArXiv:1808.04364
- 27 See A, Liu P J, Manning C D. Get to the point: summarization with pointer-generator networks. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, 2017. 1073–1083
- 28 Kingma D P, Ba J. Adam: a method for stochastic optimization. 2014. ArXiv:1412.6980
- 29 Papineni K, Roukos S, Ward T, et al. BLEU: a method for automatic evaluation of machine translation. In: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, 2002. 311–318
- 30 Lavie A, Agarwal A. METEOR: an automatic metric for MT evaluation with high levels of correlation with human judgments. In: Proceedings of the 2nd Workshop on Statistical Machine Translation, 2007. 228–231

Unsupervised paraphrasing via syntactic template sampling

Yu BAO¹, Shujian HUANG^{1,3*}, Hao ZHOU², Lei LI², Xinyu DAI¹ & Jiajun CHEN¹

1. *National Key Lab for Novel Software Technology, Nanjing University, Nanjing 210023, China;*

2. *ByteDance AI Lab, Beijing 100086, China;*

3. *Peng Cheng Laboratory, Shenzhen 518000, China*

* Corresponding author. E-mail: huangsj@nju.edu.cn

Abstract Paraphrase generation is an important natural language-processing task widely used in downstream research, such as machine translation, question answering, and text classification. Recently, some researchers have explored the advantages of paraphrasing based on syntax sampling, generating multiple paraphrases by sampling various syntactic representations from the probabilistic latent space. However, syntactic representations sampled from the posterior distribution always show high similarity and generate low-diversity paraphrases. Besides, syntactic representations sampled from the prior distribution display inferior generation quality as they can hardly match the given semantics. In this paper, we propose STVAE, introduce discrete syntactic template variables, and establish the dependency between semantic and syntactic spaces. Based on the STVAE, we further propose a two-step sampling strategy: first, we sample a discrete syntactic template from its prior distribution to pursue the diversity; then, we obtain the syntactic representation from its posterior distribution to match the semantics. Experiments show that the two-step sampling strategy can effectively combine the advantages of prior sampling and posterior sampling. The proposed STVAE achieves high diversity while keeping a competitive generation quality, thus outperforming the previous unsupervised methods.

Keywords unsupervised paraphrasing, variational autoencoder, syntax structure, sampling