



计算机科学与探索

Journal of Frontiers of Computer Science and Technology

ISSN 1673-9418, CN 11-5602/TP

## 《计算机科学与探索》网络首发论文

题目: 自动扩充关键词语义信息的诗歌生成算法  
作者: 王勇超, 周灵智, 赵亚萍, 许端清  
网络首发日期: 2022-01-19  
引用格式: 王勇超, 周灵智, 赵亚萍, 许端清. 自动扩充关键词语义信息的诗歌生成算法[J/OL]. 计算机科学与探索.  
<https://kns.cnki.net/kcms/detail/11.5602.TP.20220117.1625.002.html>



**网络首发:** 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

**出版确认:** 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

# 自动扩充关键词语义信息的诗歌生成算法

王勇超<sup>1+</sup>, 周灵智<sup>2</sup>, 赵亚萍<sup>1</sup>, 许端清<sup>2</sup>

1. 浙江大学 信息技术中心, 杭州 310027

2. 浙江大学 计算机学院, 杭州 310027

+ 通信作者 E-mail: ychwang@zju.edu.cn

**摘要:** 当前, 诗歌生成模型大多数通过用户所提供的关键词来生成符合韵律规则和音调起伏的诗歌。由于关键词蕴含的语义信息较少, 很难保证生成诗歌的质量, 容易出现上下文主题偏移的现象。针对这一问题, 本文提出了一种基于条件变分自编码器的生成模型, 该模型能够在更加丰富的语义信息指导下, 生成更符合关键字描述和用户满意度的诗歌。该模型通过采样人类创作的诗歌, 引入额外和关键词相关的语义信息, 有效估计条件变分自编码器的先验概率分布, 生成更贴合真实分布的先验概率。由于该模型自动扩充了关键词信息, 缩小了输入和输出语义信息的差距, 缓解了以往模型中普遍存在的过翻译问题。实验结果证明我们的模型无论在自动评估还是人类评估方面相比于其他模型都有更好的效果, 并成功减少了过翻译问题出现的频率, 提高了生成诗歌的流畅性。通过变化采样的范围, 成功实现了对生成诗歌写作风格的控制, 进一步证明了本文提出算法的有效性。

**关键词:** 人工智能; 自然语言处理; 自然语言生成; 诗歌生成; 条件变分自编码器

**文献标志码:** A      **中图分类号:** TP391

## Poetry generation algorithm with automatic expansion of keyword semantic information

WANG Yongchao<sup>1+</sup>, ZHOU Lingzhi<sup>2</sup>, ZHAO Yaping<sup>1</sup>, XU Duanqing<sup>2</sup>

1. Center of Information & Technology, Zhejiang University, Hangzhou 310027, China

2. Institute of Computer Science, Zhejiang University, Hangzhou 310027, China

**Abstract:** At present, most of the poetry generation models use keywords provided by users to generate poems that conform to the rules of rhythm and fluctuations in pitch. Because keywords contain less semantic information, it is difficult to guarantee the quality of generated poems, and the phenomenon of contextual theme shift is likely to occur. In response to this problem, this paper proposes a generative model based on conditional variational autoencoders, which can generate poems that are more in line with keyword descriptions and user satisfaction under the guidance of richer semantic information. By sampling human poetry and introducing additional semantic information related to keywords, the model effectively estimates the prior probability distribution of the conditional variational autoencoder, and generates a prior probability that more closely matches the true distribution. Because this model automatically expands keyword information, it narrows the gap between input and output semantic information, and alleviates the over-translation problem that is common in previous models. Experimental results show that our model has better results than other models in both automatic and human evaluation, and successfully reduces the frequency of translation problems and improves the fluency of generated poetry. By changing the range of sampling,

**基金项目:** 国家重点研发计划 (2020YFC1523101), 宁波市科技计划项目 (2021Z011), 浙江省科技项目 (2019011), 石窟寺文物数字化保护国家文物局重点科研基地支持 (浙江大学)。

This work was supported by National Key R&D Plan Project (2020YFC1523101), Ningbo Science and Technology Plan Project (2021Z011), Zhejiang Science and Technology Project (2019011) and Key Scientific Research Base of State Administration of Cultural Relics for Digital Protection of Grotto Temple Cultural Relics (Zhejiang University).

controlling the writing style of the generated poetry is successfully achieved, which further show the effectiveness of the algorithm proposed in this paper.

**Key words:** Artificial Intelligence; Natural language processing; Natural language generation; Poetry generation; Conditional Variational Autoencode

中国诗歌是中华传统文明的瑰宝。它使用简洁的语言却能够表达丰富的蕴意,成为最具代表性的中国传统文本形式之一。然而,大多数人由于缺少专业训练,缺乏在长度有限且格式受限的文本中表达丰富情感的能力。因此,研究如何利用计算机自动生成诗歌吸引了越来越多研究人员的注意。

近年来,神经网络因其强大的归纳推理能力得到了越来越广泛的应用。2014年,Zhang等人<sup>[1]</sup>首次使用了循环神经网络生成中国诗歌,获得了优于传统方法<sup>[2]</sup>的效果。后续不少研究人员针对格式控制,语句通顺,主题一致性问题,在Zhang等人<sup>[1]</sup>的模型基础上做相应的改进。为了更严格的控制诗歌的生成格式,Wang等人<sup>[3]</sup>提出使用注意力机制<sup>[4]</sup>规范生成诗歌的样式。Yan<sup>[5]</sup>等人针对语句不通顺问题,提出了迭代生成模型逐步提高生成诗歌的质量。为了解决生成诗歌无法在整首诗保持主题一致性这一问题,Wang等人<sup>[6]</sup>设计了计划框架,能够为诗歌的每一行设计一个与主题相关的子关键词,从而保证整首诗行与行之间紧密的联系。Zhang等人<sup>[7]</sup>提出用神经记忆机制有效增加模型的创造性,一定程度克服了循环神经网络倾向于学习一般范式的缺陷。受到认知心理学相关理论的启发,Yi等人<sup>[8]</sup>提出工作记忆机制,维持了有限的历史信息并选择部分信息用于当前行的生成,平衡了主题一致和创造性之间的矛盾。

2015年,Kingma等人<sup>[9]</sup>提出了一种全新的生成模型:变分自编码器。随着Bowman<sup>[10]</sup>等人证实变分自编码器的隐变量能够学习文本的全局语义信息,变分自编码器的变种条件变分自编码器

<sup>[11]</sup>(CVAE)开始在文本生成任务得到了应用。条件变分自编码器不仅能够将文本的全局语义信息表达为隐变量,而且能够在特定条件下生成对应文本。2018年,Yang等人<sup>[12]</sup>提出使用条件变分自编码器的隐变量控制诗歌每一行表达的主题,在主题一致性方面获得了优于循环神经网络的结果。与Yi<sup>[8]</sup>等人类似,Li等人<sup>[13]</sup>将对抗学习和条件变分自编码器相结合,提高了生成诗歌的新奇性和主题一致性。Yang等人<sup>[14]</sup>关注到特定情感在诗歌创作的重要地位,他们提出了半监督的条件变分自编码器用于诗歌的情感控制。Yi等人<sup>[15]</sup>将历史背景和生活经历等视为隐变量的影响因素,通过控制影响因素的设置生成不同的诗歌,有效增加了生成诗歌的多样性。

然而以上列举的所有诗歌生成模型都遵循了Zhang等人<sup>[1]</sup>对诗歌生成问题的定义:模型根据用户提供的关键词生成诗歌。关键词由有限数量的字符或词组成,一般为2-6个字符。这导致生成诗歌的字符数是关键词字符数的两倍以上。Yang<sup>[16]</sup>等人认为字符数量的巨大差距导致模型输入和模型输出的语义信息差距过大,进而给诗歌生成模型带来了过翻译问题:某些字或词会无意义的连续出现,一般为两次。然而,诗歌创作手法之一叠词的表现形式和过翻译问题的表现形式一致,无法简单的根据语法规则进行分辨。叠词可以增加诗歌的韵律性和节奏感,因此单纯在模型生成过程中加以限制字符或词的重复次数,虽然能够缓解过翻译问题,但也会导致叠词的消失,降低诗歌的美感。

为了解决这一问题,本文提出了一种新的生成模型,该模型在人类创作的诗歌中采样与关键词相关的诗歌,并利用全概率公式将采样诗歌引入到对

条件变分自编码器先验概率分布的计算中。通过将采样诗歌引入到模型生成过程,该模型自动扩充了关键词的语义信息,克服了输入和输出语义信息规模严重不匹配的问题,能在一定程度上消除过翻译问题的出现。同时,模型通过借鉴采样诗歌能够学会叠词在诗歌创作的正确使用,从而在有效避免无意义重复词出现的同时,保证了诗歌的节奏感和韵律性。Bandura<sup>[17]</sup>认为写作需要学会观察其他学习者,我们的模型以该思想为基础,通过观察人类创作诗歌的方法,能够创作出更符合人类用词习惯和美学观念的诗歌。

本文的主要工作可以概括为以下三个方面:(1)提出了新颖的诗歌生成方法,通过借鉴诗人所作的诗歌,既缓解了过翻译问题,也保留了叠词在生成诗歌中的使用。(2)利用全概率公式和蒙特卡洛方法将模仿人类诗人写诗这一抽象行为转化为对条件变分自编码器先验概率分布的求解。(3)实验证明了本文提出的方法无论在自动评估标准还是人类评估标准方面都优于其他的基础模型。

## 1 基于采样的变分自编码器诗歌生成模型

### 1.1 模型概述

我们将一首诗记作  $x$ , 把  $x_i$  视为诗歌的第  $i$  行, 从而  $x_{i,j}$  代表了一首诗的第  $i$  行第  $j$  个字符。我们使用  $c$ ,  $s$  分别表示关键词和采样诗歌。

我们的模型目标是最大化  $\ln(p_\theta(x|c))$ , 其中  $\theta$  代表  $p_\theta(x|c)$  的参数。我们通过引入隐变量  $z$  试图学习全局语义表达, 因此  $\ln(p_\theta(x|c))$  可表示为:

$$\ln(p_\theta(x|c)) = \int \ln(p_\alpha(z, x|c)) dz \quad (1)$$

遵循 kingma 等人<sup>[10]</sup>提出的方法, 我们采用变分估计方法拟合真实分布  $\ln(p_\theta(x|c))$ 。变分估计方法通过训练变分下限  $L(c, x)$  间接优化  $\ln(p_\theta(x|c))$ , 具体表现形式如下:

$$\begin{aligned} \ln(p_\theta(x|c)) &\geq L(c, x) \\ &= -KL(p_\phi(z|c, x) | q_\gamma(z|c)) \\ &\quad + \ln(p_\eta(x|c, z)) \end{aligned} \quad (2)$$

其中  $q_\gamma(z|c)$  表示隐变量  $z$  的先验概率分布,  $p_\phi(z|c, x)$  表示隐变量  $z$  的后验概率分布,  $p_\eta(x|c, z)$  可看作诗歌生成器,  $KL(\bullet)$  表示  $KL$  散度, 被用来衡量两个概率分布之间的相似度。随着模型不断的迭代训练,  $KL(\bullet)$  项变小,  $q_\gamma(z|c)$  和  $p_\phi(z|c, x)$  两者之间的差异性也越来越小。

然而,  $q_\gamma(z|c)$  只利用了关键词信息, 模型所能使用的语义信息过少, 生成诗歌容易出现过翻译问题, 即字符无意义的重复多次。扩充关键词语义信息能够缓解过翻译问题, 因此需要引入更多和关键词相关的信息。

区别于一般的条件变分自编码器, 我们使用全概率公式和蒙特卡洛方法进一步展开  $q_\gamma(z|c)$  从而引入和关键词相关的诗歌信息, 可得:

$$\begin{aligned} q_\gamma(z|c) &= \sum p(z, x|c) \\ &= \sum p(x|c) p(z|c, x) \\ &= E_{x \sim p(x|c)}(p(z|c, x)) \\ &\approx 1/k \sum_s p(z|c, s) \end{aligned} \quad (3)$$

其中,  $s$  表示根据概率分布  $p(x|c)$  从诗歌库中采样的诗歌。通过公式(3), 计算隐变量  $z$  的先验概率分布需要经历两个步骤: 1. 采样  $k$  首诗歌。2. 根据采样诗歌计算  $q_\gamma(z|c)$ 。

为了模型的训练方便, 传统的 CVAE 模型将  $q_\gamma(z|c)$  和  $p_\phi(z|c, x)$  都假设为协方差矩阵为对角矩阵的高斯概率密度函数。由于我们对 CVAE 模型的优化函数作了改动, 所以假定  $p(z|c, s)$  是协方差矩阵为对角矩阵的高斯概率密度函数, 将  $q_\gamma(z|c)$  看作高斯混合模型。我们希望找寻一个



各维度独立的高斯密度函数  $q(z|c)$ ，它和  $q_\gamma(z|c)$  之间的差异能够充分小，即它们之间的  $KL$  散度能够足够小。由于  $q(z|c)$  和  $q_\gamma(z|c)$  之间的  $KL$  散度难以计算，需要进行额外的转换从而减小计算的复杂度：

$$\begin{aligned} KL(q(z|c)|q_\gamma(z|c)) &= \\ &\int q(z|c) \ln(q(z|c)) dz \\ &- \int q(z|c) \ln(q_\gamma(z|c)) dz \\ &\leq K(c, s) \end{aligned} \quad (4)$$

$$\begin{aligned} K(c, s) &= \int q(z|c) \ln(q(z|c)) dz \\ &- \int q(z|c) \sum_s \frac{\ln(p(z|c, s))}{k} dz \end{aligned} \quad (5)$$

根据公式(4)和公式(5)，我们能够通过设置合适的参数让  $K(c, s)$  足够小，从而间接的减小  $q(z|c)$  和  $q_\gamma(z|c)$  之间的  $KL$  散度。

为了进行下一步有效推导，需要先定义一些参数。假定  $q(z|c, x)$  的期望和方差分别为  $\mu$  和  $\sigma$ ， $p(z|c, s)$  的期望和方差分别为  $\mu_s$  和  $\sigma_s$ ，隐变量  $z$  的维度表示为  $J$ ，从而公式(5)可以进一步推导为：

$$\begin{aligned} K(c, s) &= -\frac{1}{2} \sum_{j=1}^J (\ln(\sigma_j^2) - \sum_s \frac{\ln(\sigma_{s,j}^2)}{k} + 1) \\ &+ \frac{1}{2k} \sum_s \sum_{j=1}^J \left( \frac{\sigma_j^2}{\sigma_{s,j}^2} + \frac{(\mu_j - \mu_{s,j})^2}{\sigma_{s,j}^2} \right) \end{aligned} \quad (6)$$

根据公式(6)分别对  $\mu$  和  $\sigma$  求偏导得到以下结果：

$$\frac{\partial(K(c, s))}{\partial \mu_j} = -\frac{1}{k} \sum_s \frac{(\mu_j - \mu_{s,j})}{\sigma_{s,j}^2} \quad (7)$$

$$\frac{\partial(K(c, s))}{\partial \sigma_j} = -\frac{1}{\sigma_j} + \frac{1}{k} \sum_s \frac{\sigma_j}{\sigma_{s,j}^2} \quad (8)$$

将公式(7)和公式(8)置零，从而求得  $K(c, s)$  极点对应的  $\mu$  和  $\sigma$ ：

$$\mu_j = \frac{\sum_s \frac{\mu_{s,j}}{\sigma_{s,j}^2}}{\sum_s \frac{1}{\sigma_{s,j}^2}} \quad (9)$$

$$\sigma_j^2 = \frac{k}{\sum_s \frac{1}{\sigma_{s,j}^2}} \quad (10)$$

假定  $\mu_{s,j}$  的定义域为  $(a, b)$ ， $\sigma_{s,j}^2$  的定义域为  $(c, d)$ ，将公式(9)和公式(10)代入  $K(c, s)$  可求得如下不等式：

$$K(c, s) \leq \frac{1 + J(b-a)^2 d - J \ln(c)}{2} \quad (11)$$

因此，如果我们设置适当的超参数  $J, a, b, c, d$ ，可以获得一个高斯分布  $q(z|c)$ ，它和目标高斯分布  $q_\gamma(z|c)$  的  $KL$  散度小于一个较小的常数上限。

模型的主要框架如图1所示。我们把模型分为三个部分：先验部分，后验部分和生成器。模型在先验部分通过关键词采样更接近用户需求的诗歌(采样过程如图2所示)，并通过两者计算隐向量的先验概率分布。后验部分使用关键词和相应诗歌计算隐变量的后验概率分布。先验概率分布和后验概率分布在训练过程中差异变小。生成器通过隐变量(训练时为后验部分隐变量，测试时为先验部分隐变量)、关键词生成目标诗歌。

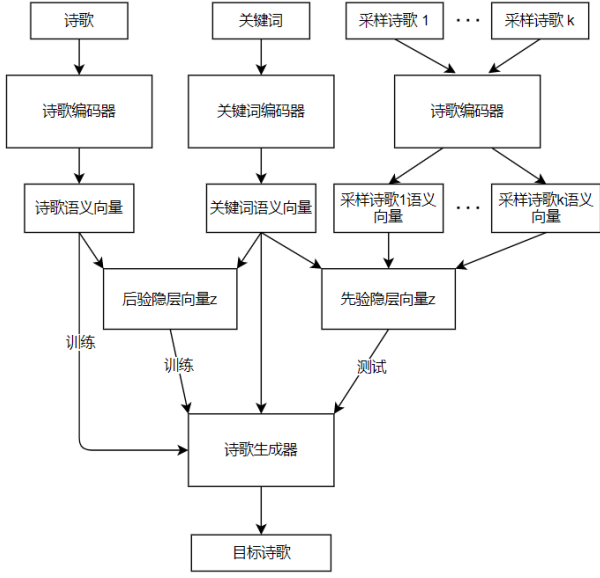


图1 基于采样的条件变分自编码器框架

Fig.1 Sampling-based conditional variational autoencoder framework

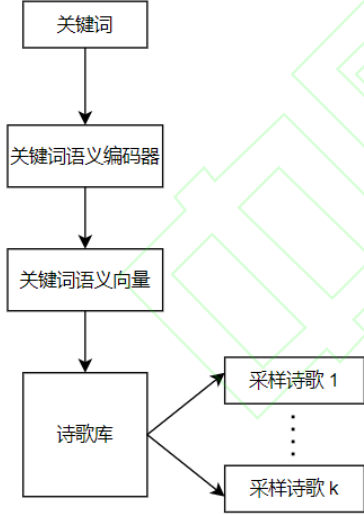


图2 基于关键词的采样过程

Fig.2 Keyword-based sampling process

## 1.2 先验概率计算

诗歌生成任务的 CVAE 模型通常使用循环神经网络学习关键词的语义表达，并将语义表达作为单层的全连接神经网络的输入，进而计算先验概率分布。本文提出的模型根据采样诗歌计算先验概率分布，与传统的先验概率分布计算方式存在较大差异。为了采样人类创作的诗歌，我们

需要预训练去噪自编码器模型。我们采用了两个双向递归神经网络（分别记作  $G_p$  和  $G_c$ ），将  $G_p$  和  $G_c$  的前向隐向量和后向隐向量连接，从而得到  $h_p$  和  $h_c$ ， $h_p$  和  $h_c$  各自对应诗歌和关键词的语义表达。然后连接  $h_p$  和  $h_c$  并加入随机噪声，作为生成器的输入。当去噪自编码器模型训练完成后，我们使用关键词语义向量  $h_c$  帮助采样诗歌。

任意一首人类创作诗歌都对应着一个关键词。我们存储了诗歌库中所有诗歌对应的关键词语义向量  $h_c$ ，并将其记作  $H_c$ 。模型通过关键词  $c$  对应的语义向量  $h_c$ ，进而计算概率分布  $p(x|c)$ ：

$$p(x|c) = \text{softmax}(h_c * H_c^T) \quad (12)$$

我们根据概率分布  $p(x|c)$  采样  $k$  首诗歌，使用  $G_p$  和  $G_c$  得到  $k$  个  $h_p^r$  和  $k$  个  $h_c^r$ ，记作  $h_{p,1}^r \dots h_{p,k}^r$  和  $h_{c,1}^r \dots h_{c,k}^r$ 。需要注意的是，我们固定  $G_p$  和  $G_c$  的参数，不再随着训练而更新。这是因为  $H_c$  必须在训练过程中保持不变，从而保证基于相同条件采样的诗歌都遵循一致的概率分布。遵循 Li 等人<sup>[12]</sup>的方式，我们使用激活函数为  $\tanh(\bullet)$  的单层全连接神经网络生成  $p(z|c, a, x_s)$  的  $\mu_{s,i}$  和  $\sigma_{s,i}$ ：

$$\begin{bmatrix} \mu_{s,i} \\ \sigma_{s,i} \end{bmatrix} = \text{MLP} \left( \begin{bmatrix} h_{p,i}^r \\ h_{c,i}^r \end{bmatrix} \right) \quad (13)$$

模型根据公式(9)和公式(10)计算先验分布  $q(z|c, x)$  的  $\mu_{prior}$  和  $\sigma_{prior}$ ，然后使用重参数技巧采样隐向量  $z$ 。

## 1.3 后验概率计算

计算后验概率分布时，依旧采用上文提及的  $G_p$  和  $G_c$  计算， $G_p$  和  $G_c$  的参数在训练过程固定。我们使用关键词和相应诗歌计算  $h_p'$  和  $h_c'$ 。模型使用线性变换将  $h_p'$  和  $h_c'$  的连接映射到相应的  $\mu_{post}$

和  $\mu_{post}$  :

$$\begin{bmatrix} \mu_{post} \\ \sigma_{post} \end{bmatrix} = MLP\left(\begin{bmatrix} h_p^t \\ h_c^t \end{bmatrix}\right) \quad (14)$$

与计算先验分布类似, 我们使用同样的方法采样符合后验分布的隐向量  $z$ 。

## 1.4 生成器

我们的生成器结构参考了 Yan<sup>[4]</sup>等人的工作, 并在其基础上做了部分的改动。我们使用 GRU 作为基础的生成器, 相比 RNN 它能更好的传输语义。具体计算过程如下:

$$s_{i,j} = GRU(s_{i,j-1}, [e(x_{i,j-1}); g_{i-1}]) \quad (15)$$

$$s_{i,0} = f(h_c, z, o_i) \quad (16)$$

$$p(x_{i,j} | x_{i,1...j-1}, x_{1...i-1}, w) = \text{softmax}(f(s_{i,j})) \quad (17)$$

其中  $x_{i,j}$  表示诗歌第  $i$  行第  $j$  个字符,  $e(\bullet)$  表示嵌入向量,  $;$  表示向量的连接运算,  $o_i$  是控制当前生成行长度的独热向量,  $[1,0]$  表示行长度为 5,  $[0,1]$  表示行长度为 7,  $g_i$  表示上下文语义向量。

我们需要进一步说明如何计算  $g_i$ 。对  $g_i$  的计算类似于卷积神经网络, 我们定义窗口大小为  $d$ , 因此  $g_i$  的计算可表示为:

$$a_{i,t} = f([s_{i,t}; \dots; s_{i,t+d-1}]) \quad (18)$$

$$g_i = f(g_{i-1}, \sum_t a_{i,t}), g_0 = 0 \quad (19)$$

## 2 实验设置与结果

### 2.1 数据集

我们的实验主要在 THU Chinese Classical Poetry Corpus<sup>[18]</sup>(THU-CCPC)上进行。THU-CCPC 被分成训练, 验证和测试三个数据集, 且每首诗都有作者, 朝代及相应关键词信息。THU-CCPC

数据集情况如表 1 所示。

表 1 数据集概况

Tab.1 Data set overview

	诗歌数	行数
训练集	97147	388588
验证集	28287	112748
测试集	8767	35068

### 2.2 基线模型

为了更好的探索模型的性能, 我们需要选择一些基础模型和本文所提出的模型进行对比。这些模型需要有一定程度的代表性且和我们模型有一定的关联性。因此, 我们选择了以下几种模型开展对比:

Seq2seq。一种传统的可用于文本生成任务的模型, 我们所参考的 Yan<sup>[6]</sup>的模型可被认为属于 seq2seq 模型。

Seq2seq+mem 模型。Zhang 等人<sup>[7]</sup>提出的一种模型, 该模型使用神经记忆机制试图克服 seq2seq 模型缺乏创造力的缺点, 他们将一些诗歌存储起来帮助模型生成诗歌, 该模型的思想和我们提出的模型具有一定的相似性, 和该模型的对比是必要的。

CVAE 模型。CVAE 模型近几年来被广泛应用于文本生成任务, 九歌系统大量使用了该模型<sup>[14,15,18]</sup>。我们的模型可看作 CVAE 的变种, 需要通过对比实验研究本文在 CVAE 模型上做的改变所带来的优缺点。

CVAE+mem 模型。我们将 mem 机制运用于 CVAE 模型, 从而更好地比较 mem 机制和基于采样的 CVAE 模型的优劣。

### 2.3 评价标准

为了公正的评价生成的诗歌, 我们需要使用一些评价标准:

首先我们使用了机器翻译指标 BLEU<sup>[19]</sup>。BLEU 评价指标被广泛应用于机器翻译领域,用来衡量真实结果和机器翻译结果的差异。由于诗歌生成和机器翻译存在一定的相似性,之前的很多工作也大量使用 BLEU 来评价生成诗歌的质量。

其次我们采用了叠词相似度 (Repetition similarity) 来进行评价。我们模型的目的是为了减少叠词和无意义重复之间的矛盾,因此我们需要定义相关的评价标准。我们分别统计测试集和模型生成诗歌从每个位置开始出现连续两个字符的频率,并把每个位置当作一个维度,从而得到两个向量表达。我们计算两个向量的余弦相似度,衡量真实结果和生成诗歌之间的差异。我们假定真实结果不存在无意义的重复,因此余弦相似度越大越符合我们的目标。

最后我们采用了人类评估的方式来予以评价。诗歌创作是人类创造力的体现,完全采用自动评估标准评价诗歌质量存在缺陷。为了更好的衡量模型之间的差异,我们邀请诗歌领域的 10 名专家来评估生成诗歌的质量。出于公平的考虑,他们并不知道评价的诗歌由哪个模型生成。我们遵循之前相关工作的定义,要求专家从四个方面评价诗歌:一致性,流利度,意义和诗意(具体细节可见于表 2)。每个方面的打分范围都为 0-5,分数越高表示效果越好(打分值可取小数点后一位)。我们将 10 名专家在每个标准的打分分别求均值得到最终的人工评估结果。

表 2 人类评估标准

Tab.2 Human Evaluation Standard

标准	描述
一致性	一首诗是否表现出一致的主题
流利度	一首诗在语法上是否满足定义
意义	一首诗的内容是否有意义
诗意	一首诗是否具有诗的属性

## 2.4 自动评估结果分析

表 3 记录了不同模型生成诗歌的自动评估结果。值得一提的是,我们将本文所提出的模型命名为 S-CVAE (条件变分编码器的采样版本)。

由于 CVAE 在语义表征方面优于 seq2seq 模型,因此相较于 seq2seq 和 seq2seq+mem 在 BLEU 指标上表现得更好。我们不再讨论 S-CVAE 和 seq2seq 模型以及 seq2seq+mem 在 BLEU 分数的对比,因为 S-CVAE 从某种意义上来说从属于 CVAE 模型。S-CVAE 相较于 CVAE 在 BLEU 指标略有优势,这是因为 S-CVAE 能够减少无意义重复词的产生,从而提高诗歌的生成质量。我们发现 S-CVAE 在 BLEU 指标分数最高,在叠词相似度指标上却比 CVAE+mem 和 seq2seq+mem 表现得要差。因此我们重点讨论 S-CVAE 和 mem 机制的对比。

表 3 自动评估标准模型对比

Tab.3 Automatic evaluation of standard model comparison

	BLEU	Repetition similarity
Seq2seq	6.39	0.485
Seq2seq+mem	6.23	0.507
CVAE	8.89	0.513
CVAE+mem	8.43	0.522
S-CVAE	9.23	0.521

Mem 机制在部分思路上和我们的模型相似,我们试图探索 mem 机制和 S-CVAE 对生成诗歌的不同影响。我们通过观察实验结果作出以下推论:

(1) mem 机制能够缓解过翻译问题。根据表 3 的实验结果,我们发现无论是将 mem 机制与 seq2seq 模型或 CVAE 模型结合,都能够显著的提升叠词相似度指标。Mem 机制根据人类创作诗歌在每个位置的词频调整诗歌的生成,从而让生成诗歌符合一般的用词规律。因此,mem 机制既能缓解过翻译问题,也能维持叠词的使用。



(2) mem 机制会破坏诗歌的生成质量。从表 3 中我们发现 seq2seq+mem 在 BLEU 分数上比 seq2seq 模型还要低, 是 BLEU 分数最低的基准模型。mem 机制通过诗歌每个位置的词频调整生成字符的概率分布, 从而减少了无意义重复字符的产生。然而 mem 机制无法保证所参考的诗歌和用户提供关键词有关, 甚至存在参考诗歌和用户提供关键词完全相背的情况。当 mem 机制根据完全无关的诗歌调整目标诗歌的生成, 不可避免的会让生成诗歌和关键词匹配度下降, 从而降低了 BLEU 分数。

(3) S-CVAE 在缓解过翻译问题方面略逊于 mem 机制。表 3 显示 S-CVAE 在叠词相似度指标高于 CVAE 模型, 符合我们对该模型的预期。与 CVAE 相比, 我们的模型在输入信息中增加了采样诗歌, 大大增加了输入的语义信息规模, 从而克服了导致过翻译问题的直接原因, 减少了无意义重复字符的产生。然而, S-CVAE 先将采样诗歌和用户提供关键词压缩成隐向量, 再借由 LSTM 神经网络解压成生成诗歌, 在压缩和解压的过程中不可避免的带来了信息丢失的问题。而 mem 机制直接将 softmax 层计算的生成词概率分布通过诗歌库对应位置的词频进行调整, 保留了更多的信息, 能够更好的抑制无意义重复词的产生。

(4) S-CVAE 能够在缓解过翻译问题的同时, 保证诗歌的流畅性。我们可以把 S-CVAE 看作 mem 机制的泛化形式。mem 机制直接通过词频调整诗歌的生成, 可能带来了完全与当前关键词无关的信息, 干扰了诗歌的生成质量。我们的模型先通过语义相似程度对诗歌库中的诗歌做了初步过滤, 并利用上述所推导的公式进一步调整了采样诗歌对生成过程的影响程度, 从而让生成诗

歌与关键词语义一致。mem 机制强行让生成词的概率函数接近诗歌库中的词频, 破坏了生成字符和关键词语义的匹配程度。而 S-CVAE 对生成字符的调整较为平滑, 能够在一定程度上纠正 CVAE 模型的过翻译问题, 并能够保证生成诗歌的流畅性。

## 2.5 人工评估结果分析

表 4 记录了不同模型生成诗歌的人工评估结果。人工评估的结果和自动评估的结果高度一致, S-CVAE 在一致性, 流利度两个指标明显高于其他的基线模型, 这是因为 S-CVAE 会事先通过关键词采样最相关的诗歌, 通过采样诗歌指导诗歌生成过程, 和之前的模型相比能够提取更多的语义, 更好的在诗歌生成过程中保持语义的一致性。并且, S-CVAE 克服了输入信息和输出信息语义规模的巨大差距并从采样诗歌中学习到了更丰富的语义信息, 从而缓解了过翻译问题。虽然 mem 机制能够用更粗暴的方法减少无意义重复词的产生, 然而它也在缓解过翻译问题的同时, 引入了大量无关的信息甚至有害的信息, 破坏了生成诗歌的流利度以及一致性。人工评估的结果再一次佐证了我们对 mem 机制和 S-CVAE 之间的分析。图 3 和图 4 展示了 S-CVAE 模型的生成结果, 从直观上展现了模型在一致性和流利度取得的进展。

表 4 人工评估标准模型对比

Tab.4 Human evaluation of standard model comparison

	一致性	流利度	意义	诗意
Seq2seq	2.94	2.46	2.25	2.67
Seq2seq+mem	3.08	2.61	2.17	2.61
CVAE	3.28	2.77	2.60	3.19
CVAE+mem	3.35	2.88	2.57	3.17
S-CVAE	3.68	3.30	2.58	3.12

小阁疏帘映晚霞，白云深处野人家。

西风吹起梧桐树，独立斜阳数点鸦。

图 3 S-CVAE 七言古诗生成示例

Fig.3 Seven-character ancient poem generation example of S-CVAE

袅袅垂杨柳，柔条拂晚风。

美人肠断处，知在画桥东。

图 4 S-CVAE 五言古诗生成示例

Fig.4 Five-character ancient poem generation example of S-CVAE

## 2.6 风格化分析

为了进一步探索采样对生成诗歌的影响，我们对不同诗人创作的诗歌采样，比较生成诗歌风格之间的差异性。以李白和杜甫为例，模型使用相同关键词分别采样李白和杜甫的诗歌。统计李白和杜甫诗作中 TF-IDF 最高的 20 个字符，分别记作集合  $A_{\text{李白}}$  和  $A_{\text{杜甫}}$ 。对生成诗歌执行同样的操作，记作  $B_{\text{李白}}$  和  $B_{\text{杜甫}}$ 。我们使用 Jaccard 系数( $|A \cap B|/|A \cup B|$ )衡量生成诗歌与真实诗歌风格之间的相似性，实验结果如表 5 所示。表 5 证实了采样范围的不同会影响生成诗歌的风格。

表 5 Jaccard 系数

Tab.5 Jaccard coefficient

	$B_{\text{李白}}$	$B_{\text{杜甫}}$
$A_{\text{李白}}$	0.25	0.14
$A_{\text{杜甫}}$	0.11	0.29

使用图 5 和图 6 两个具体示例进一步佐证采样策略的不同对生成诗歌风格的影响：

白鹭洲边柳，秋风拂钓舟。

相逢休问讯，明月满江楼。

图 5 带有李白风格的生成古诗

Fig.5 Generative ancient poems with Li Bai style

白鹭洲边柳，秋风拂钓舟。

故人今夜月，肠断楚江秋。

图 6 带有杜甫风格的生成古诗

Fig.6 Generative ancient poems with Du Fu style

图 5 和图 6 所示古诗都以明月作为关键词，却表现出不同的意境。图 5 的诗歌采样李白诗作，表现出洒脱浪漫的情怀，而图 6 的诗歌采样杜甫诗作，表达了哀愁忧伤的情感。

## 3 结束语

在本文中，我们提出了一种基于采样的条件变分自编码器算法。它能够缓解过翻译问题，减少无意义重复词的产生，从而生成质量更高的诗歌，自动评估结果证明了方法的有效性。同时，我们根据自动评估结果进一步分析了 S-CVAE 模型和 mem 机制对生成诗歌过程的不同影响，人工评价标准也佐证了分析的准确。风格化分析进一步显示了采样对于诗歌生成的巨大影响，通过采样不同诗人的作品，能够在相同关键词的条件下表现出不同的风格。然而，S-CVAE 在意义和诗意这两个方面仍有进步空间。这是因为 S-CVAE 虽然采样了和关键词匹配的诗歌，保证了生成诗歌和关键词的高度一致性，却在一定程度上忽略了诗歌和诗歌之间表达语义方式不匹配的问题。部分诗歌表达情感的方式较为含蓄，部分诗歌表

达情感的方式比较直白, 导致生成诗歌的表达方式出现割裂。表达方式的割裂让生成诗歌的理解难度有了一定的提高, 且无法充分体现诗歌中心思想的层层递进, 造成意义表达不够明确, 同样的问题也出现在 mem 机制中。未来, 我们可以结合半监督的条件变分自编码器用于诗歌的情感表达控制, 从而保证采样诗歌表达方式的一致性, 进一步完善 S-CVAE 在意义和诗意存在的不足, 生成蕴意更为丰富的诗歌。

## 参考文献:

- [1] ZHANG X, LAPATA M. Chinese poetry generation with recurrent neural networks[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2014, pages 670–680.
- [2] HE J, ZHOU M, and JIANG L. Generating Chinese Classical Poems with Statistical Machine Translation Models[C]//Proceedings of the 26th AAAI Conference on Artificial Intelligence, 2012, pages 1650–1656.
- [3] WANG Q, LUO T, WANG D, et al. Chinese song iambics generation with neural attention-based model[C]//Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, 2016, pages 2943–2949.
- [4] BAHDANAU D, CHO K and BENGIO Y. Neural machine translation by jointly learning to align and translate[J]. arXiv preprint arXiv:1409.0473, 2014.
- [5] YAN R, JIANG H, LAPATA M, et al. I, poet: automatic chinese poetry composition through a generative summarization framework under constrained optimization[C]//Proceedings of the 23rd International Joint Conference on Artificial Intelligence, 2013 pages 2197–2203.
- [6] WANG Z, HE W, WU H, et al. Chinese poetry generation with planning based neural network[C]//Proceedings of the 26th International Conference on Computational Linguistics, 2016, pages 1051–1060.
- [7] ZHANG J, FENG Y, WANG D, et al. Flexible and creative chinese poetry generation using neural memory[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, 2017, pages 1364–1373.
- [8] YI X, SUN M, LI R, et al. Chinese poetry generation with a working memory model[C]//Proceedings of the 28th International Joint Conference on Artificial Intelligence,, 2018, 4553–4559.
- [9] KINGMA D, WELLING M. Auto-encoding variational bayes[J]//arXiv preprint arXiv:1312.6114, 2013.
- [10] BOWMAN S, VILNIS L, VINYALS O, et al. Generating sentences from a continuous space.[C]//The SIGNLL Conference on Computational Natural Language Learning, 2016.
- [11] YAN X, YANG J, SOHN K, et al. Attribute2image: Conditional image generation from visual attributes[C]// European Conference on Computer Vision, 2016, pages 776–791.
- [12] YANG X, LIN X, SUO S, et al. Generating thematic chinese poetry using conditional variational autoencoders with hybrid decoders[C]// Proceedings of the 28th International Joint Conference on Artificial Intelligence, 2018.
- [13] LI J, SONG Y, ZHANG H. et al Generating classical chinese poems via conditional variational autoencoder and adversarial training[C]// Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2018, 3890–39.
- [14] YANG C, SUN M, YI X, et al. Stylistic chinese poetry generation via unsupervised style disentanglement[C]// Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2018, 3960–3969.
- [15] YI X, LI R, YANG C, et al. MixPoet: Diverse Poetry Generation via Learning Controllable Mixed Latent Space[C]// Proceedings of the 34th AAAI Conference on Artificial Intelligence, 2020, pages 9450–9457.
- [16] YANG Z, CAI P, FENG Y, et al. Generating Classical Chinese Poems from Vernacular Chinese[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2019.
- [17] BANDURA B. Social cognitive theory: An agentic perspective[M]//Annual Review of Psychology, 2001, 52(1):1–26.
- [18] GUO Z, YI X, SUN M, et al. Jiuge: A Human-Machine Collaborative Chinese Classical Poetry Generation System[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: System Demonstrations, 2019.
- [19] PAPINENI S. Blue: A method for Automatic Evaluation of Machine Translation[C]//Meeting of the Association for Computational Linguistics, 2002.



王勇超(1975-), 男, 浙江杭州, 硕士, 高工, 主要研究方向为计算机视觉、自然语言处理。

WANG Yongchao, born in 1975, master, senior engineer. His research interest include computer vision and natural language processing.



周灵智(1997-), 男, 江西鹰潭, 硕士研究生, 主要研究方向为自然语言处理、知识生成。

ZHOU Lingzhi, born in 1998, master's degree student. His research interest include natural language processing and knowledge generation.



赵亚萍(1975-), 女, 浙江杭州, 硕士, 高工, 主要研究方向为知识图谱。

ZHAO Yaping, born in 1975, master, senior engineer. Her research interest include knowledge graph.



许端清(1966-), 男, 浙江杭州, 博士, 教授, 博士生导师, 主要研究方向为人工智能、机器学习。

XU Duanqing, born in 1966, Ph.D., professor, doctoral supervisor. His research interest include artificial intelligence and machine learning.