

更具有感情色彩的诗歌生成模型^①



廖荣凡, 沈希忠, 刘 爽

(上海应用技术大学 电子信息工程系, 上海 201418)

通讯作者: 沈希忠, E-mail: xzshen@yeah.net

摘 要: 五言绝句是我们传统文学的宝藏, 给人独特的语言美感和审美体验. 使用机器生成绝句诗歌对机器理解人类语言有着积极的探索意义. 依据诗歌语言自身韵律和对仗特点, 我们在诗歌数据集和对联数据集上联合训练诗歌生成模型. 模型包括语义模型和文字规则模型, 语义模型创新性地使用一维卷积网络提取诗歌文字的语义特征, 学习诗歌语义的主题信息. 文字规则模型使用带注意力机制的编码解码器, 学习诗歌文字的对仗特征. 实验结果表明模型可以很好地生成符合诗歌规则, 表现诗人情感的诗句, 如“感时花溅泪, 愁路竹林心. 秋风草树色, 夜雨寒风声.”

关键词: 五言律诗; 语义模型; 文字规则模型; 一维卷积; 注意力机制

引用格式: 廖荣凡, 沈希忠, 刘爽. 更具有感情色彩的诗歌生成模型. 计算机系统应用, 2020, 29(5): 46-51. <http://www.c-s-a.org.cn/1003-3254/7366.html>

Poetry Generation Model with More Emotional Information

LIAO Rong-Fan, SHEN Xi-Zhong, LIU Shuang

(Department of Electronic Information Engineering, Shanghai Institute of Technology, Shanghai 201418, China)

Abstract: The five-character quatrain is the treasure of Chinese traditional literature, rendering people a unique language aesthetic and aesthetic experience. The process of machine-generated quatrain has a positive exploration of machine acquisition on human language. Inspired by the rhythm and antithesis features of poetry language itself, we trained language memory model on poetry datasets and couplet datasets. The model consists of a semantic model and a textual rule model. The semantic model uses a one-dimensional convolution network to extract the semantic features of poetry and learn the semantic information of the poetry. The word model uses an encoder-decoder model with attention mechanism to learn the isocolon features of poetry writing. The experimental results show the language memory model can generate poems that conform to the rules of poetry and our aesthetics.

Key words: five-word quatrain; semantic model; text rule model; one-dimensional convolution; attention mechanism

诗歌是一种优美的语言形式, 诗以其意象组合和语言韵律给人带来独特的审美体验. 五言绝句是我国传统文学的结晶有着很高的文学和艺术价值. 一方面绝句诗歌本身有严格的对仗和格律要求, 诗人写作时应仔细推敲每一个字的使用, 以最精炼的文字表达出最细腻的情感, 另一方面这种结构规则含义丰富的诗歌形式, 为计算机理解人类语言提供了良好的范本. 近年来有越来越多的学者开始使用计算机模拟诗歌的规

则和诗人的写作方式来生成主题明确、语言通畅的绝句诗. 这些工作对于计算机理解人类语言和思维都有着积极的探索意义.

五言律诗最基本的格律要求是: 重音位置的平仄相对. 平仄律是六朝及隋唐诗人根据当时的语言形态和审美观念不断探索而形成的诗体规范^[1]. 绝句中基本的对仗要求是: 诗歌上下两句对应位置动词对动词、名词对名词. 古代诗人可以凭借对格律的运用和

① 收稿时间: 2019-09-03; 修改时间: 2019-10-08; 采用时间: 2019-10-24; csa 在线出版时间: 2020-05-07

文字的把控即兴赋诗而浑然天成,如李白《将进酒》“呼儿将出换美酒,与尔共消万古愁”,曹植七步成诗“本是同根生,相煎何太急”。这对于计算机自动生成诗歌的启发是我们可以更加关注诗歌语言文字本身的搭配关系,在诗句之间加入对联的规则,以增强诗歌语言的韵律和感情色彩。

目前将文字向量化的算法可以捕捉到文字间细腻的语法和语义关系^[2,3],含义相近的字在向量空间中的分布也更相近。词向量的表达方便了句子建模,Zhang等^[4]使用卷积神经网络提取语句在不同层次上的语义和模式特征,进行句子间的语义匹配。这可以看作是使用卷积神经网络建模句子的语义表征。循环神经网络在语音分析和文本翻译上的表现证明它可以很好的提取和记忆输入信息前后之间的联系^[5]。

受到二维卷积神经网络在图像处理应用中的启发,我们创新性地将一维卷积网络应用到绝句诗歌的生成当中,提出了一种更具有感情色彩的诗歌生成模型。

我们的诗歌生成模型由语义模型和文字规则模型构成。语义模型根据上句(以句号分句,逗号分隔的子句为半句)的输入预测下半句的输出,以增加诗歌主题和上下文的连贯性。文字规则模型根据语义模型输出的前半句,预测输出的后半句,使诗句文字在感情色彩上和主题一致。语义模型和文字规则模型共同完成一首诗歌的生成。

具体的语义模型由一维卷积网络和双向循环网络构成,一维卷积网络用于提取诗句的抽象语义特征,双向循环网络用于记忆语义特征之间的结构关系;文字规则模型由词嵌入网络和带注意力机制的编码解码器构成,用于学习诗歌文字的对仗关系和文字在高维空间的结构关系。

1 相关工作

关于诗歌自动生成的研究已经有了几十年的历史,早期的诗歌生成基于规则和模板^[6,7]之后基于统计机器学习算法的模型被用到了诗歌生成当中如SMT (Statistical Machine Translation)^[8],文献^[9]设计了一种基于短语的统计机器翻译模型,以前一句诗歌生成后一句诗歌。文献^[10]提出以统计机器翻译为理论基础,将格律诗的上下句关系映射为统计翻译模型中源语言与目标语言的关系设计了融入诗词领域知识的统计机器翻译模型用于诗歌生成。当前神经网络算法在诗歌

自动生成任务中应用广泛,体现出了明显的优势。现在大部分诗歌生成基于循环神经网络RNN (Recurrent Neural Network)。Zhang等^[4]最早将单一的循环神经网络RNN应用到绝句诗的生成当中,Zhang等在词嵌入的基础上使用语句卷积模型提取上下文特征向量,捕捉诗歌的上下文信息,然后使用RNN依据上下文向量和之前生成的诗句得出下一句诗。但是早期的RNN和他们使用的卷积模型较为简单,为了提升语言的连贯性他们还引入了额外的SML (Statistical Machine Learning) 特征和N-gram语言模型。随后Yan等^[11]等提出基于RNN的编码解码模型ipoet,采用循环迭代的方式使模型每一次的输出作为下一次的输入,模仿诗人创作反复打磨的过程。Wang等^[12]采用Planning Model对诗歌的主题进行分割,诗歌的每一行对应一个子主题,然后诗句在子主题的指引下由编码解码网络自动生成,作者还调整了编码网络的结构使编码器可以编码子主题和之前生成的诗句,以生成主题连贯的诗歌。文献^[13]中,作者将诗歌生成问题看成序列到序列的学习问题,使用带注意力机制的编码解码模型学习字符间的联系和输入输出诗句间的相关性,并采用了词生成句,句生成句,段落生成句3种独立的诗歌生成方式。文献^[14]作者认为神经网络的概率本质虽然可以学习到诗歌中出现频率较高的语言模式生成语句通顺的诗歌但它严重忽视了诗歌创新性的美学意义,进而提出了记忆增强的神经模型(memory-augmented neural model)结合带注意力机制的神经网络和诗歌特征映射模型生成有一定创新性的诗歌。文献^[15]作者提出诗歌生成的动态记忆模型分别独立保存写作的主题向量作为主要信息,主题向量会随着诗句的生成进行更新,在生成每一行诗句时模型读取最相关的主题向量指导生成当前诗句。此外作者还引用了风格嵌入控制来生成语言风格不同的诗歌。

上述的诗歌自动生成模型多把重点放在诗歌的主题控制与文字的流畅性上,我们的模型将重点放在诗歌语言文字的本身上,以获得语言的流畅性、主题的一致性和更强的感情色彩。

2 模型

诗歌生成模型由语义模型和文字规则模型构成,在诗歌语料库和对联语料库上联合完成训练。模型可以根据输入的文字描述生成诗歌,但为方便对比不同

模型生成的诗句, 实验中采用 50 条给定诗歌的首半句 (以“,”分隔半句;“.”分隔一句) 作为输入。

诗歌生成方法如图 1 所示, 给定半句诗歌 S_{11} 作为输入, 由文字规则模型 WM (Word Model) 生成半句诗 S_{12} 。 S_{11}, S_{12} 为模型生成诗歌的第一句。然后将 S_{11}, S_{12} 输入语义模型 SM (Semantic Model) 再生成半句诗 S_{21} 。之后将 S_{21} 作为下一句诗的给定输入, 生成 S_{22} 。完整的诗歌有两句组成“ $S_{11}, S_{12}, S_{21}, S_{22}$ ”。

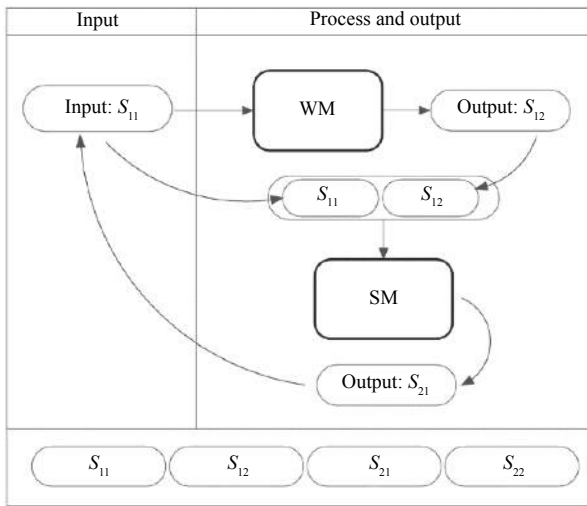


图 1 模型方法图

在生成一句诗歌的过程中, 语义模型 SM 在给定诗句 (上文) C 后, 输出下文 m 个汉字 W , 记 $W = \{w_1, w_2, \dots, w_m\}$, 形式化的表示为:

$$W = SM(C) \quad (1)$$

文字规则模型 WM 根据给定的汉字 W , 计算生成下文 m 个字的概率, 记 $S = \{w_{m+1}, w_{m+2}, \dots, w_{m+3}\}$ 为这 m 个字的特定排列, 形式化表达为:

$$P(S) = WM(W) \quad (2)$$

模型整体可以看作在给定条件下, 计算生成诗句 S 的概率, 如式 (3)。给定的条件为: 写作描述 C 和模型生成的前 m 个字。

$$P(S|w; C) = \prod_{i=1}^m P(w_{i+m}|w_{1:i+m-1}, C) \quad (3)$$

诗歌生成的问题转换为最大化这个条件概率的问题。

2.1 语义模型

2.1.1 诗歌数据处理

首先对包含 30 000 首五言诗的数据集进行字频统计, 统计每一个使用到的汉字出现的次数。考虑到古

汉语和现代汉语的不同, 为生成更通俗易懂的诗句, 我们过滤掉频次少于 15 字和所有包含这些字的诗歌, 保留 18 000 首诗歌和 2379 个汉字用于模型训练。然后建立字典, 将保留的每一个汉字映射到一个整数。最后构建训练数据集 $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$, 其中 $x_i \in R^{12}, (1 \leq i \leq N)$ 表示每一首诗每一句的整数映射向量, $y_i \in R^6, (1 \leq i \leq N)$ 表示对应于 x_i 所表示诗句之后的半句诗歌的整数映射向量。例如, 处理“白日依山尽, 黄河入海流。”查询每一个字对应的整数索引, 编码诗句为 $[12, 24, 43, 1785, 456, 1, 45, 79, 55, 28, 56, 2]$, 则相应的 $x_i = [12, 24, 43, 1785, 456, 1, 45, 79, 55, 28, 56, 2]^T$, $y_i = [45, 79, 55, 28, 56, 2]^T$, N 表示训练数据集的总样本数。

2.1.2 模型结构

使用 Tensorflow1.13.0 中的 Keras API 搭建神经网络模型。模型的输入 $x \in R^{12}$ 为 12 维的列向量, 经过两次步长为 1, 卷积核为 $(3, \cdot)$ 的一维卷积网络处理后将输入向量 x 转化为特征张量 $T \in R^{8 \times 128}$ 用于表征诗句的语义信息。每次一维卷积的数学描述为:

$$T_{i,j}^C = \sigma \left(\sum_{k=1}^K T_{i+k-1}^l \cdot f_k^j \right) \quad (4)$$

其中, $\sigma(\cdot)$ 表示非线性函数, $T_{i,j}^C$ 表示卷积特征矩阵 T^C 的第 i 行, 第 j 列元素; $T_i^l \in R^{1 \times N}$ 表示输入张量第 i 行的 N 个元素; $f_k^j \in R^{1 \times N}$ 表示第 j 个卷积核第 k 行的 N 个元素, 这里 $f \in R^{K \times N}$ 。

将卷积后的特征张量 T^C 输入双向 LSTM (Long Short-Time Memory) 模型^[16]记忆学习诗歌抽象语义特征。LSTM 的数学描述^[17]可表达为:

$$h_t = f_0 \circ \tanh(c_t) \quad (5)$$

$$f_0 = \sigma(w_0 \cdot [h_{t-1}; x_t] + b_0) \quad (6)$$

$$c_t = f_t \circ c_{t-1} + f_u \circ \tilde{c}_t \quad (7)$$

$$\tilde{c}_t = \tanh(w_c \cdot [h_{t-1}; x_t] + b_c) \quad (8)$$

$$f_u = \sigma(w_u \cdot [h_{t-1}; x_t] + b_u) \quad (9)$$

$$f_f = \sigma(w_f \cdot [h_{t-1}; x_t] + b_f) \quad (10)$$

其中, $\sigma(\cdot)$ 表示非线性函数, 符号 \circ 表示向量对应元素间的乘法, h_t 表示输出向量, c_t 表示记忆向量 x_t 表示输入向量, w 表示权重矩阵, b 表示偏置向量。

双向 LSTM 将数据按时间步的顺序过程和倒序过程分别输入到两个 LSTM 模型中, 最后将两个模型的

输出向量 h_t 拼接在一起。

LSTM 网络后连接两层全连接层,用于预测模型的输出 $\hat{y} \in R^{1 \times m}$,语言记忆网络的损失函数为均方误差函数:

$$L = \frac{1}{2m} \sum_{i=1}^m (y_i - \hat{y}_i)^2 \quad (11)$$

2.2 文字规则模型

2.2.1 对联数据处理

针对对联数据集构建字典,将数据集中的每一个字映射到一个特定的整数,构建训练数据集 $T = \{(x_1, y_1, t_1), (x_2, y_2, t_2), \dots, (x_N, y_N, t_N)\}$ 其中 $x_i (1 \leq i \leq N)$ 表示每一首对联上联汉字的整数映射向量, $y_i (1 \leq i \leq N)$ 表示每一首对联下联汉字的整数映射向量, $t_i (1 \leq i \leq N)$ 由 y_i 的字符串开头加入一个特定标记字符形成, N 表示训练数据集的总样本数。

2.2.2 模型结构

模型先使用词嵌入层将输入的文字序列映射到高维的向量空间中,拓展文字序列向量的表征能力。

将对联字典中汉字对应的整数编码 w_i 使用 Word2Vector 方法嵌入到 E 维的向量空间中。令 V 表示词典的大小, $L \in R^{E \times |V|}$ 表示字符的词向量矩阵,它的列对应着字符嵌入的空间维度。 $e(w_k) \in R^{|V|}$, $e_i = 0, i \neq k$, $e_k = 1$, e_i 表示 $e(w_k)$ 中的各元素。即 $e(w_k)$ 代表第 k 个字符的独热向量表示,则输入 m 个字符的序列 w_1, w_2, \dots, w_m 经过词嵌入后可以表示为:

$$(T^e)^T = [L \cdot e(w_1), L \cdot e(w_2), \dots, L \cdot e(w_m)] \quad (12)$$

其中, T^e 表示 m 个字符的嵌入张量, $(T^e)^T \in R^{E \times m}$, $T^e \in R^{m \times E}$ 。

对输入的词嵌入张量,模型采用带注意力机制的序列到序列结构,由编码器(EN_{LSTM})对输入张量编码提取特征向量后由解码器(DE_{LSTM})结合注意力机制预测输出张量。其运算过程可以描述为^[18]:

编码器接受上联输入的词嵌入张量 x_t 和上一时间步编码器(EN_{LSTM})的隐藏层输出 h_{t-1} ,计算当前时间点的隐藏层输出 h_t 。如式(13):

$$h_t = EN_{LSTM}(x_t, h_{t-1}) \quad (13)$$

解码器接受下联输入的词嵌入张量 y_t 和上一时间步解码器(DE_{LSTM})的隐藏层输出 s_{t-1} ,计算当前时间点的隐藏层输出 s_t ,如式(14):

$$s_t = DE_{LSTM}(y_t, s_{t-1}) \quad (14)$$

由编码器和解码器的隐藏层输出 h_t 、 s_t 计算一个权重分数和编码器每一个时间步隐藏层输出的权重,如式(15)、式(16):

$$e_{ij} = \text{score}(s_i, h_j) = h_j^T s_i \quad (15)$$

$$\alpha_{ij} = \text{Softmax}(e_{ij}) \quad (16)$$

由编码器每一个时间步隐藏层输出的加权平均生成上下文向量 c_i 。

$$c_i = \sum_{j=0}^n \alpha_{ij} h_j \quad (17)$$

将上下文向量 c_t 与解码器隐藏层的输出 s_t 拼接起来计算模型输出的概率:

$$P(y_t | y_{1:t-1}, x) = \text{Softmax}(W_s [c_t; s_t]) \quad (18)$$

模型的目标函数为交叉熵损失函数:

$$J_t = \sum_{(x,y) \in T} -\log p(y|x) \quad (19)$$

其中, T 为训练数据集。

3 实验

模型在 30 000 首五言诗歌数据集和 50 000 首对联数据集上进行训练。我们使用 3 个模型进行诗歌生成的比较,一是类似 Yan 等^[11]的 LSTM 模型;二是类似 Wang 等^[19]的注意力模型,三是我们的诗歌生成模型。3 个模型分别在相同测试集上生成 50 首诗歌,用于测评。

3.1 评价标准

参考文献[2]我们使用 BLEU^[20]作为诗歌生成模型的评估方法。这种方法在评估时需要给出参考文本,对于如何获取参考文本,He 等^[8]的思想是如果两句诗有相同的主题,那这两句诗的后两句诗可以相互参考。我们模型的诗歌生成基于给定诗歌第一句的前半句,所以自然地我们选择给定诗句的全诗作为 BLEU 评分的参考文本。

因为诗歌这种特殊形式的文本其美学价值和给人的情感体验有较强的主观性,所以我们也设计了人工评判环节,为使评判有较强的客观性,选择 4 条评判标准: 1) 对仗,模型生成诗句间的词语是不是相互对仗; 2) 韵律,模型生成的诗歌是不是按五言律诗的规则平仄相对; 3) 意象,诗歌意象是诗歌情感意义的主要载

体,这里考察诗歌中是否包含常见的文学意象,如春风、秋月、旧竹、寒梅等;4)连贯性,诗歌整体上下句表达的情感是否流畅或一致.每条标准按1~5分评分.

3.2 评估结果

使用二元 BLEU 对上述 3 个模型生成诗句评分,每个模型根据给定的 50 条输入生成 50 首诗歌,50 首诗歌的平均 BLEU 得分如表 1 所示.

表 1 模型 BLEU

模型	得分
双层 LSTM	0.2773
注意力模型	0.2984
本文模型	0.3024

每个模型 BLEU 得分前 5 的诗句分别由表 2, 表 3, 表 4 所示.

表 2 双层 LSTM 模型

诗歌生成	得分
国破山河在, 年山人条身. 有弦时人旧, 衣欢烟声晚.	0.3247
岭外音书断, 边道春叶未. 寄骑桥见处, 何我入心深.	0.3107
感时花溅泪, 钟中得水年. 我一林又似, 相明云青看.	0.2950
绝代有佳人, 深坐觉明间. 带日幽城手, 花无长松一.	0.2950
闲门向山路, 水门往若朝. 明起看结人, 霜朝营事望.	0.2946

表 3 注意力模型

诗歌生成	得分
大漠孤烟直, 长城满水流. 秋景复秋风, 春风复夜阑.	0.3745
浮云游子意, 清月入人心. 秋风入画屏, 夜雨湿衣裳.	0.3608
荷风送香气, 竹气入云阴. 花阴入水流, 日暮向山阴.	0.3562
风鸣两岸叶, 日暮九江花. 春风万里来, 万里一山水.	0.3414
松月生夜凉, 风波入晚凉. 日暮向秋风, 秋风入夜阑.	0.3347

表 4 本文模型

诗歌生成	得分
千山鸟飞绝, 万里出花中. 初且价可喜, 更有故人心.	0.3744
大漠孤烟直, 长城满水流. 客赌木三在, 人心一万年.	0.3665
随山将万转, 应与一年春. 前奴艳客君, 明月有人心.	0.3573
细草微风岸, 长门落日寒. 有锦看不如, 无处有无声.	0.3489
山随平野尽, 春意有心深. 日晃百花色, 风波万里流.	0.3454

由实验结果可以发现虽然双层 LSTM 模型可以提取诗歌基本的结构特征,但它生成的诗歌在 BLEU 评分中明显不如注意力模型和本文模型,而且主题不明显.这里需要指出的是,双层 LSTM 模型只在五言诗歌的数据集上进行训练,而后两个模型在诗歌数据集和对联数据集上联合训练,这种对比可能对双层 LSTM

模型有些不公平,但是在尽我们努力所翻阅的文献中,前人的诗歌生成工作只利用了诗歌数据集,而没有加入对联数据集来加强学习诗歌文字间的对仗关系,所以这里对于基线模型双层 LSTM,我们只使用了诗歌数据集对其进行训练.

BLEU 评分考察更多的是文字之间的相关性,有较高得分的诗句不一定有诗歌的美感,明显的主题和鲜明的情感色彩,所以我们根据上文提到的诗歌评分标准,邀请 8 位来自文学专业的同学对每个模型生成的 50 首诗歌打分.每个模型生成诗歌平均得分情况如表 5 所示.

表 5 人工评估得分

模型	对仗	韵律	意象	连贯性	平均
双层 LSTM	2	2.5	2.5	2	2.25
注意力模型	3.5	3.25	3.5	3.25	3.375
本文模型	3.75	3.25	3.75	3.5	3.56

3.3 实验分析

在诗歌和对联数据集上训练的注意力模型和本文的模型可以较好的表现诗歌的对仗和韵律,一方面是因为注意力机制可以更好的关注诗歌对应位置的字符关系,另一方面是因为对联本身有较强的韵律,字词之间也有明显的对应关系.如注意力模型生成的诗句“大漠孤烟直, 长城满水流. 秋景复秋风, 春风复夜阑.”,“大漠”对“长城”,“孤烟直”对“满水流”,“秋景”对“春风”,“秋风”对“夜阑”,明显的对仗关系和相似的意象组合可以较好地表现出诗歌的美感和韵律.又如“饮马渡秋水, 幽林出晚风. 秋景湿人衣, 春风入画屏.”、“海上生明月, 江中出晚风. 江水向秋波, 山阴复夜色”,但注意力模型的缺点是一方面很多相似的意象组合限制了诗歌主题的丰富性,如秋景和春风,秋波和夜色,明月和晚风常常相对;一方面不相似的意象组合也使诗歌主题损失了整体性,如上句“大漠长城”,下句“秋景春风”.意境的破坏同时也弱化了诗歌的情感表达.

本文模型使用一维卷积神经网络将上句诗歌的 12 个字符依次输入到网络中预测下文的输出,将文字生成看成特殊的时间序列预测问题.我们相信一维卷积网络的卷积核可以像词嵌入网络一样提取语言的语义结构并结合 LSTM 模型构成具有一定语言记忆功能的模型.如在生成的诗歌“树树皆秋色, 花花亦未央. 离惜堤竹旧, 春风草未长.”中,上句“未央”与下句“离惜”相呼应使诗歌主题更一致,更有美感,同时也使诗歌有

了更鲜明的情感色彩又如:“迟日江山丽, 秋风草树闲. 听晴司可喜, 看子向陵阡。”上句“日”与下句“晴”相呼应;“浮云游子意, 清月入人心. 初晴弥可喜, 更与故人心。”上句“清月”与下句“初晴”相呼应, 上句“游子意”与下句“故人心”相呼应通过实验的对比分析, 我们发现诗歌和对联数据集上联合训练的模型可以较好地提升生成诗歌对仗和韵律的效果, 同时也具有更强的语言表现力和诗歌的美感. 我们的诗歌生成模型在提高诗歌生成的对仗和韵律效果的同时也改善了主题生成的一致性. 一致的主题, 流畅的语言和对仗的意象使模型生成的诗歌更具有更强得感情色彩.

4 总结

我们的语义模型创新性地提出使用一维卷积网络和长短时记忆网络 (LSTM) 提取诗句的语义信息, 增强诗句之间主题的一致性和情感的相关性. 文字规则模型为带注意力机制的编码解码模型, 用于学习诗歌文字的韵律和字词的对仗关系. 模型在诗歌和对联数据集上联合训练. 实验结果表明我们的模型取得了预期的效果, 可以较好地呈现出诗歌的语言特征和情感色彩, 使机器生成的诗歌有类似诗人即兴赋诗的美感和表现力.

参考文献

- 李飞跃. 中国古典诗歌平仄律的形成与嬗变. 中国社会科学, 2015, (3): 136–157.
- Pennington J, Socher R, Manning C. Glove: Global vectors for word representation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Doha, Qatar. 2014. 1532–1543.
- Mikolov T, Chen K, Corrado G, *et al.* Efficient estimation of word representations in vector space. arXiv: 1301.3781, 2013.
- Zhang XX, Lapata M. Chinese poetry generation with recurrent neural networks. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Doha, Qatar. 2014. 670–680.
- LeCun Y, Bengio Y, Hinton G. Deep learning. Nature, 2015, 521(7553): 436–444. [doi: [10.1038/nature14539](https://doi.org/10.1038/nature14539)]
- Gervás P. An expert system for the composition of formal Spanish poetry. In: Macintosh A, Moulton M, Coenen F, eds. Applications and Innovations in Intelligent Systems VIII. London: Springer, 2001. 19–32.
- Wu XF, Tosa N, Nakatsu R. New hitch haiku: An interactive Renku poem composition supporting tool applied for sightseeing navigation system. Proceedings of the 8th International Conference on Entertainment Computing. Berlin, Heidelberg, Germany. 2009. 191–196.
- He J, Zhou M, Jiang L. Generating Chinese classical poems with statistical machine translation models. Proceedings of the 26th AAAI Conference on Artificial Intelligence. Toronto, Canada. 2012. 1650–1656.
- 何晶, 周明, 蒋龙. 基于统计的汉语格律诗生成研究. 中文信息学报, 2010, 24(2): 96–103. [doi: [10.3969/j.issn.1003-0077.2010.02.013](https://doi.org/10.3969/j.issn.1003-0077.2010.02.013)]
- 蒋锐澄, 崔磊, 何晶, 等. 基于主题模型和统计机器翻译方法的中文格律诗自动生成. 计算机学报, 2015, 38(12): 2426–2436.
- Yan R. i, Poet: Automatic Poetry Composition through recurrent neural networks with iterative polishing schema. Proceedings of the 25th International Joint Conference on Artificial Intelligence. New York, NY, USA. 2016. 2238–2244.
- Wang Z, He W, Wu H, *et al.* Chinese poetry generation with planning based neural network. arXiv: 1610.09889, 2016.
- Yi XY, Li RY, Sun MS. Generating Chinese classical poems with RNN encoder-decoder. Proceedings of the 16th Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data. Nanjing, China. 2017. 211–223.
- Zhang JY, Feng Y, Wang D, *et al.* Flexible and creative Chinese poetry generation using neural memory. arXiv: 1705.03773, 2017.
- Yi XY, Sun MS, Li RY, *et al.* Chinese poetry generation with a working memory model. arXiv: 1809.04306, 2018.
- Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation, 1997, 9(8): 1735–1780. [doi: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735)]
- Greff K, Srivastava RK, Koutník J, *et al.* LSTM: A search space odyssey. IEEE Transactions on Neural Networks and Learning Systems, 2017, 28(10): 2222–2232. [doi: [10.1109/TNNLS.2016.2582924](https://doi.org/10.1109/TNNLS.2016.2582924)]
- Luong MT, Pham H, Manning CD. Effective approaches to attention-based neural machine translation. arXiv: 1508.04025, 2015.
- Wang QX, Luo TY, Wang D, *et al.* Chinese song iambics generation with neural attention-based model. arXiv: 1604.06274, 2016.
- Papineni K, Roukos S, Ward T, *et al.* BLEU: A method for automatic evaluation of machine translation. Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. Philadelphia, PA, USA. 2002. 311–318.