

图像如何激发诗歌:从图像中产生中国古典诗歌 使用记忆网络

Linli Xu [†], Liang Jiang [†], Chuan Qin [†], Zhe Wang [‡], Dongfang Du [†]

[†]安徽省大数据分析与应用重点实验室,
中国科学技术大学计算机科学与技术学院

[‡]蚂蚁金服集团人工智能部
linlixu@ustc.edu.cn, jal@mail.ustc.edu.cn, chuanqin0426@gmail.com
wz143459@antfin.com, dfdu@mail.ustc.edu.cn

抽象的

随着神经模型和自然语言处理的最新进展,文言文的自动生成

诗歌因其艺术性和艺术性而受到广泛关注
文化价值。以前的工作主要集中在生成
诗歌给定关键字或其他文本信息,而视觉
诗歌的灵感很少被探索。生成
从图像中写诗比从文本中生成诗更具挑战性,因为图像包含非常丰富的视
觉
用几个关键词无法完整描述的信息,一首好诗应该传达形象

准确。在本文中,我们提出了一种基于记忆的神经模型,该模型利用图像生
成诗歌。具体来说,提出了一种带有主题记忆网络的编码器-解码器模型来
生成中国古典诗歌

图片。据我们所知,这是第一部作品
试图从图像中生成中国古典诗歌
与神经网络。结合人类评价和定量分析的综合实验研究表明,所提出的模
型可以生成诗歌

准确地传达图像。

介绍

中国古典诗词是中国文化中宝贵而重要的一期。在超过

2000年在中国,写了数百万首诗词来赞美英雄人物,美丽

风景、爱情等。中国古典诗歌至今仍以其简洁的结构、韵律的美感和

丰富的情感。文言文有不同的流派
诗歌,包括唐诗、宋词、清诗等,各有不同的结构和规律。

其中,四行诗最为流行,四行,每行五七个字。

四行诗的行遵循特定的规则,包括
有规律的节奏模式,其中最后一个字符
第一(可选),第二和第四行必须属于
相同的节奏类别。另外,每个汉字都是
与一个音调相关联,该音调可以是 Ping (电平音)
或 Ze (向下的音调),并且四行诗需要遵循预先定义的音调模式,该模式调节
音调
不同位置的字符 (Wang 2002)。一个例子

版权所有 c 2018,人工促进协会
情报 (www.aaii.org)。版权所有。

望庐山瀑布
Cataract on Mount Lu
李白
Li Bai
日照香炉生紫烟, (*ZPPZZP)
The sunlit Censer peak exhales a wreath of cloud,
遥看瀑布挂前川。 (*P*ZZPP)
Like an upended stream the cataract sounds loud.
飞流直下三千尺, (*P*ZPPZ)
Its torrent dashes down three thousand feet from high,
疑是银河落九天。 (*ZPPZZP)
As if the Silver River fell from azure sky.

表 1:7 字符四行诗的示例。每个人的基调
字符显示在每行的末尾,其中“P”代表 Ping (水平音),“Z”代表 Ze (向下

音),“*”表示音调都可以。押韵
字符带有下划线。

中国著名古典诗人的绝句
李白如表1所示。

中国古典诗歌严格的韵律和声调规则对自动生成中国诗歌提出了重大挑战。
近年来,各种尝试

从关键词等文本信息中自动生成古诗词。之中

他们,基于规则的方法 (Tosa,Obar 和 Minoh 2008;
Wu,Tosa 和 Nakatsu 2009;Netzer 等人 2009;Oliveira
2009; 2012), 遗传算法 (Manurung 2004; Zhou,
你和丁 2010; Manurung,里奇和汤普森
2012)和统计机器翻译方法 (Jiang and
周 2008; He, Zhou, and Jiang 2012) 已经发展起来。
最近,随着深度神经
网络,一些基于诗歌生成算法
已经提出了关于神经网络的范式
序列到序列的学习,生成诗歌
一行一行,每一行都是通过取前一行生成的
行作为输入 (Zhang 和 Lapata 2014;Yi,Li 和 Sun 2016;
王等人。 2016a; 2016b;张等人。 2017)。但是,以前的作品存在
限制,包括主题漂移和
仅在第一行考虑用户的写作意图引起的语义不一致。此外,

有限数量的具有正确顺序的关键字通常是
需要 (Wang et al. 2016b),这限制了

过程。

另一方面,视觉灵感更自然
比写诗的文本更直观。人们可以写
表达审美或感伤的诗歌
壮丽的山脉和湍急的河流等引人入胜的风景映入眼帘。结果,有

通常一首诗和一幅图像之间存在显式或隐式的对应关系,一首诗要么描述一个
场景,或给读者留下视觉印象。例如,表 1 中的诗代表了一个图像

壮观的瀑布从高山上下落。由于这种内在的相关性,从图像生成中国古典
诗歌成为一个有趣的研究课题。

据我们所知,图像的视觉灵感有
在经典的自动生成中很少被探索
中国诗歌。从图像生成诗歌的任务
通常比从关键词生成诗歌更具挑战性,因为图像中包含非常丰富的视觉
信息,这需要复杂的表示作为传达基本视觉特征和

图像的语义概念到诗歌生成器。在
此外,要生成一首与图像一致且本身连贯的诗歌,应仔细考虑主题流

在生成的字符序列中进行操作。

在本文中,我们提出了一个 Encoder-Decoder 框架
用主题记忆生成中国古典诗歌
图像,它以提取的关键词的形式将直接的视觉信息与语义主题信息相结合
从图像。我们将诗歌生成视为一个序列到序列的学习问题,其中生成一首
诗
逐行生成,每一行都是通过考虑所有先前的行来生成的。此外,我们引入
一个内存
网络支持无限数量的关键词提取
从图像中动态确定生成诗歌中每个字符的潜在主题。这解决了主题漂移
和语义不一致的问题,同时解决了需要有限数量的关键词的限制

在以前的作品中以适当的顺序。同时,为了利用不包含在语义概念中的图
像的视觉信息,我们将直接视觉信息整合到

提出的编码器-解码器框架,以确保图像和诗歌之间的对应关系。实验

结果表明,我们的模型在
利用图像中的视觉和主题信息,它
可以生成传达图像的高质量诗歌
准确和一致。
本文的主要贡献是:

- 1. 我们考虑生成经典的新研究课题
意象汉诗,不仅是为了娱乐或教育,更是一种探索

自然语言处理和计算机视觉。

- 2. 我们使用记忆网络来解决
主题漂移和语义不一致,同时解决
限制需要有限数量的关键词
以前作品中的正确顺序。
- 3. 我们整合了概括语义主题的关键词和
直接的视觉信息来捕捉信息

在生成诗歌时融入图像。

相关工作

在过去,诗歌创作一直是一项具有挑战性的任务
几十年。已经提出了多种方法,大多数
其中专注于从文本中生成诗歌。之中
他们提出了基于短语搜索的方法 (Tosa,
小原和箕面 2008; Wu,Tosa 和 Nakatsu 2009)
日本诗歌一代。语义和语法模板用于 (Oliveira 2009),而遗传算法

受雇于 (Manurung 2004;Manurung,Ritchie 和
汤普森 2012;周、尤和丁 2010)。在工作中
of (Jiang and Zhou 2008; Zhou, You, and Ding 2010;
He, Zhou, and Jiang 2012),诗歌生成被视为
统计机器翻译问题,下一行在哪里
是通过翻译上一行生成的。(Yan et al. 2013) 中的另一种方法通过总
结来生成诗歌
用户的查询。

最近,深度神经网络已被应用
在自动诗歌生成中。基于 RNN 的框架
在 (Zhang and Lapata 2014)中提出,其中每首诗
线是通过将先前生成的线作为
输入。在 (Yi, Li, and Sun 2016)的工作中,关注
机制被引入诗歌生成,其中
提出了基于注意力的编码器-解码器模型来顺序生成诗句。不一样的古典
流派
中国诗歌产生于 (Wang et al. 2016a),即
第一个生成中国歌曲抑扬格的作品,每个
可变长度的行。然而,上面介绍的神经方法都有一个局限性,即一个人的
写作意图
用户仅在第一行被考虑在内,其中
将导致以下几行中的主题漂移。为了解决这个问题,
(Wang et al. 2016b) 中提出了一种改进的基于注意力的编码器-解码器
模型,该模型为每个诗行分配一个子主题关键词。因此,数量

关键词固定为诗行数,关键词必须手动排序,

限制了方法的灵活性。

在这项工作中,我们解决了以前工作中的问题
通过引入具有无限关键字容量的记忆网络,可以动态确定主题

对于每个字符。记忆网络是 (Weston,Chopra 和 Bordes)提出的一类
神经网络
2014)用记忆组件来增强循环神经网络来模拟长期记忆。在工作中

(Sukhbaatar et al. 2015),记忆网络得到了扩展
具有端到端的培训机制,这使得
模型更普遍适用。张等人。(2017) 将记忆网络引入自动诗歌生成

首次在写新诗的同时利用现有诗歌中的先验知识。在测试阶段,

通过使用记忆网络,一些现有的诗歌被表示为提供先验知识的外部记忆

为新诗。与 (Zhang et al. 2017) 相比,我们以完全不同的方式使用记忆
网络,其中
我们将关键字表示为记忆实体,这样我们就可以
无限制地处理关键字并动态确定每个字符的主题。

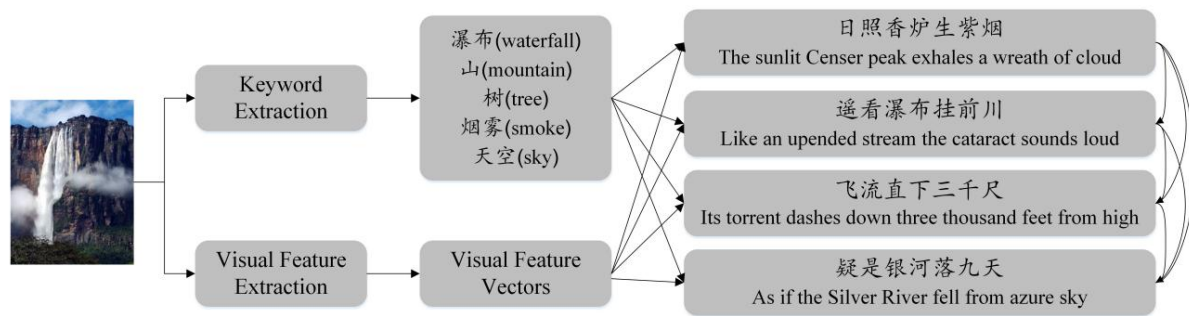


图 1:使用基于内存的图像到诗歌生成器从图像生成诗歌的流程示意图。

基于记忆的图像到诗歌生成器

图像将丰富的信息从视觉信号传递到可以激发好诗的语义主题。建立一个自然的

图像和诗歌之间的对应关系,我们提出了一个

该框架集成了概括图像重要主题的语义关键字,以及直接的

生成一首诗的视觉信息。具体来说,给定一个

图像,包含语义主题的关键字被提取为

生成诗歌时的大纲,同时利用视觉信息来体现诗歌未传达的信息

关键字。

框架

给定一个图像 I ,我们正在生成一首诗 P ,它由 L 首诗行 $\{l_1, l_2, \dots, l_L\}$ 组成。在框架中作为

如图1所示,我们首先提取一组关键词

$K = \{k_1, k_2, \dots, k_N\}$ 和一组视觉特征向量

$V = \{v_1, v_2, \dots, v_B\}$ 来自 I ,带有关键字提取器和

基于卷积神经网络 (CNN) 的视觉特征

分别提取器。这首诗然后由

线。具体来说,在生成第 i 行 l_i 时,之前生成的行用 $l_{1:i-1}$ 表示,它是从 l_1 到 l_{i-1} 的连接,关键字 K 和视觉特征向量 V 在基于内存

Image to Poem Generator (MIPG) 模型,这是关键框架中的组件。

如图 2 所示,MIPG 模型本质上是

由两个模块组成的编码器-解码器模型:

基于图像的编码器 (I-Enc) 和基于内存的解码器

(M-12 月)。在 I-Enc 中,视觉特征 V 是从

带有 CNN 的图像 I ,而双向门控循环

单元 (Bi-GRU) 模型 (Cho et al. 2014) 用于构建

来自先前生成的行 $l_{1:i-1}$ 的语义特征 H 。

在 M-Dec 中,第 i 首诗行,由一系列字符 $\{y_1, \dots, y_G\}$ 表示,是基于关键

字 K 生成的,

视觉特征向量 V 以及语义 fea

从前几行中提取 H 。生成每个字符

$y_t \in l_i$,我们首先用注意力机制将 V 和 H 转换为 y_t 的动态表示,然后是

主题记忆网络旨在动态确定

y_t 的主题。最后,使用增强图像之间一致性的主题偏差概率能力预测 y_t

和诗。

基于图像的编码器 (I-Enc)

在 I-Enc 中,如图 2 的下半部分所示,我们

首先将图像 I 编码为 B 个局部视觉特征向量

$V = \{v_1, v_2, \dots, v_B\}$,每个都是 D_v 维

表示对应于图像的不同部分。我们使用 CNN 模型作为视觉特征提取器。

具体来说, V 是通过获取 CNN 特征提取器的某个卷积层的输出来生成的

以我为输入

$$V = \text{CNN}(I).$$

同时,我们使用 Bi-GRU 对前面的行进行编码

生成的诗歌 $l_{1:i-1}$ 的形式,其形式为字符嵌入 $\{x_1, x_2, \dots, x_C\}$ 的序列

对应的隐藏向量 $H = [h_1, h_2, \dots, h_C]$,

其中 C 表示 $l_{1:i-1}$ 的长度, h_j 是前向隐藏向量的串联

向量 \overleftarrow{h}_j 在 Bi-GRU 的第 j 步。那是,

$$\overrightarrow{h}_j = \text{GRU}(\overrightarrow{h}_{j-1}, x_j),$$

$$\overleftarrow{h}_j = \text{GRU}(\overleftarrow{h}_{j+1}, x_j),$$

$$h_j = [\overrightarrow{h}_j; \overleftarrow{h}_j].$$

提取的视觉特征 V 和语义特征 H

然后在 M-Dec 中利用 I-Enc 来动态确定

图像的哪些部分和前面的哪些内容

生成每个字符时应该关注的行。

基于内存的解码器 (M-Dec)

在 M-Dec 中,如图 2 的上半部分所示,每个

line $l_i = \{y_1, \dots, y_G\}$ 是逐个字符生成的。

具体来说,在第 t 步,我们使用另一个 GRU

维护一个内部状态 st 来预测 y_t 。 st 已更新

基于 $st-1$, y_t-1 , H 和 V 循环,并且可以表示为

$$st = f(st-1, y_t-1, t, \overrightarrow{h}_t),$$

$$H_t = \text{attH}(H),$$

$$v^{\wedge}t = \text{attV}(V),$$

其中 f 表示更新内部状态的函数

GRU.attH 和 attV 表示注意力的功能

将 H 和 V 转换为动态表示 h^{\wedge}

分别表示图像的哪些部分和什么

h^{\wedge} 和 $v^{\wedge}t$

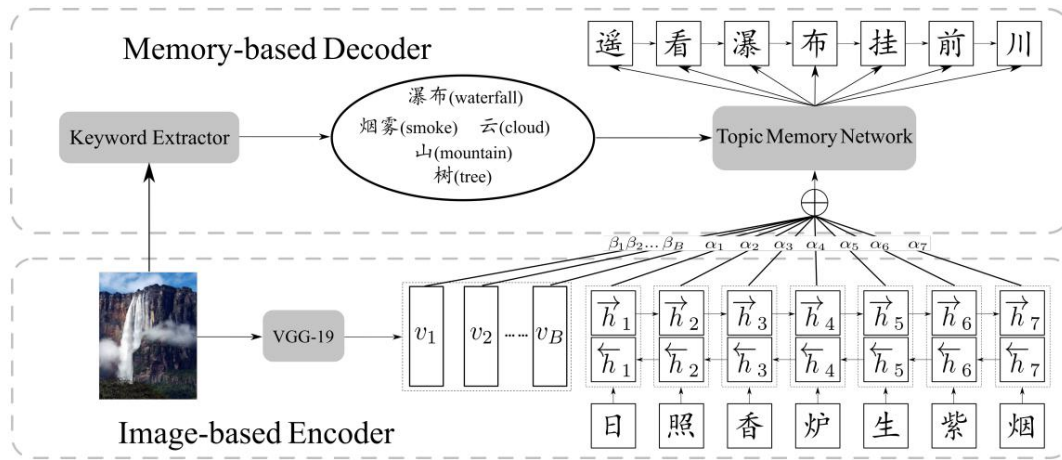


图 2:基于记忆的图像到诗歌生成器 (MIPG) 的图示。

在生成下一个字符时,模型应该关注前面几行中的内容。更正式地说,

$$H_t = \text{attH}(H) = X \alpha_{tj} h_j, h_j \in H,$$

其中 α_{tj} 是注意力模型计算的第 j 个字符的权重:

$$\alpha_{tj} = \frac{\exp(\alpha_{tj})}{\sum_{n=1}^C \exp(\alpha_{tn})},$$

$$\alpha_{tn} = u \cdot \tanh(W \alpha_{tj} - 1 + U a_n),$$

其中 α_{tj} 是第 t 步对 h_j 的注意力得分。 u 、 W 和 U 是要学习的参数。

类似地,使用以 V 作为输入的 attV 计算 $v^{\wedge t}$:

$$v^{\wedge t} = \text{attV}(V) = X \beta_{tj} v_j, v_j \in V,$$

$$\beta_{tj} = \frac{\exp(\beta_{tj})}{\sum_{n=1}^B \exp(\beta_{tn})},$$

$$\beta_{tn} = u \cdot \tanh(W \beta_{tj} - 1 + U b_n).$$

我们没有直接使用 s_t 来预测第 t 个字符 y_t ,而是引入了一个主题记忆网络,通过将 s_t 作为输入并输出一个主题感知状态向量 o_t ,它动态地为 y_t 确定一个合适的主题,它不仅包含图像和前面的行,也是生成 y_t 的潜在主题。然后使用多层感知器从 o_t 预测 y_t 。

主题记忆网络。考虑到图像所传达的丰富信息,如果关键字太少,基本上不可能完全描述一张图像。同时,话题漂移问题会削弱图像与诗歌之间的一致性。为了解决这些问题,

如图 3 所示,我们使用主题记忆网络,其中每个记忆实体都是从图像中提取的关键字,通过考虑从图像中提取的所有关键字来动态确定每个字符的潜在主题。

为此,我们使用 Clarifai¹ 提供的 general-v1.3 模型提取一组关键字 $K = \{k_1, \dots, k_N\}$,并将 K 中的每个关键字 k_j 编码为两个记忆向量*输入记忆向量 q_j 用于计算 k_j 对 y_t 的重要性,输出记忆向量 m_j 包含 k_j 的语义信息。

具体来说,对于每个具有 C_j 字符的关键字 $k_j \in K$,我们将 k_j 编码为语义向量。考虑到 C_j 字符的顺序,我们使用 Bi-GRU 将 k_j 编码为前向隐藏状态向量序列 $[-\rightarrow q_1, \dots, -\rightarrow q_{C_j}]$ 和后向隐藏状态向量序列 $[\leftarrow q_1, \dots, \leftarrow q_{C_j}]$ 。然后通过连接最后一个前向状态 $-\rightarrow q_{C_j}$ 和第一个后向状态 $\leftarrow q_1$ 来计算输入记忆向量 q_j ,即 $q_j = [-\rightarrow q_{C_j}; \leftarrow q_1]$ 。

每个关键字 k_j 的输出内存表示 m_j 由 k_j 中所有字符的嵌入向量的平均值计算,

$$m_j = \frac{1}{C_j} \sum_{n=1}^{C_j} e_{jn},$$

其中 e_{jn} 表示在训练期间学习到的 k_j 中第 n 个字符的词嵌入。

使用第 t 步的隐藏状态向量 s_t ,我们根据 s_t 与关键字的输入记忆表示 q_j 之间的相似性计算每个关键字 $k_j \in K$ 的重要性 z_j 。这可以表述如下:

$$z_j = \text{softmax}(s_t \cdot q_j), 1 \leq j \leq N.$$

给定关键字 $[z_1, \dots, z_N]$ 的权重, y_t 的潜在主题向量 d_t 由

¹<https://clarifai.com>

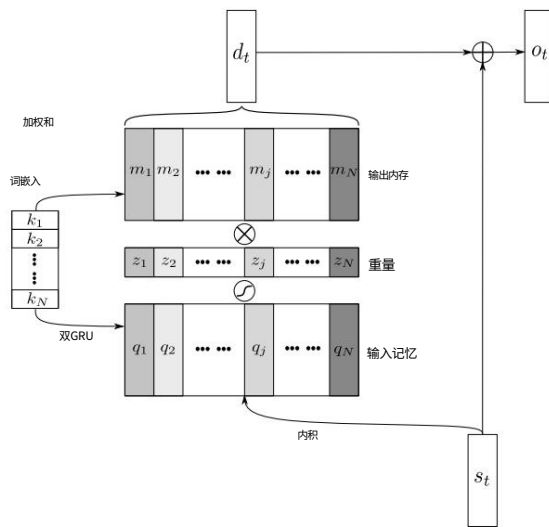


图 3: 主题记忆网络示意图。

关键字 $[m_1, \dots, m_N]$ 的输出内存表示

$$d_t = \sum_{j=1}^N z_j m_j$$

基于 d_t 和 s_t ,我们计算一个主题感知状态向量 $o_t = d_t + s_t$ 以在预测 y_t 时整合来自先前生成的字符的潜在主题、视觉特征和语义信息。

最后,基于 o_t 、 $v^{\wedge}t$ 、 h^{\wedge} 和 y_{t-1} 预测第 t 个字符 y_t 。此外,为了增强诗歌和图像偏见的丰富生成,我们从通用词汇表EG中提取的关键词中的字符。具体来说,我们定义了一个通用词汇表EG,其中包含所有可能的字符,以及一个包含 K 中所有字符的主题词汇表ET,满足 $ET \subseteq EG$ 。为了预测 y_t ,除了通用字符概率 $p_G(y_t)$ 之外,我们通过 $p_G(y_t = w) = g_G(o_t, v^{\wedge}t, h^{\wedge}t)$, $w \in EG$ 计算主题字符概率 $p_T(y_t)$,

$$p_T(y_t = w) = (g_T(o_t, v^{\wedge}t, h^{\wedge}t), w \in EG \setminus ET,$$

其中 g_T 和 g_G 表示对应于多层感知器的函数,后跟 softmax,分别计算 p_T 和 p_G 。 p_T 和 p_G 相加以计算字符概率 p ,并选择具有最大概率的字符作为下一个字符:

$$p(y_t = w) = \lambda p_T(y_t = w) + p_G(y_t = w), p(y_t = \arg \max_{w \in EG} p(y_t = w)), w \in EG, y_t =$$

其中 λ 是平衡通用概率 p_G 和主题偏差概率 p_T 的超参数。

实验

数据集

在本文中,我们感兴趣的是在给定图像的情况下生成 qua trains 的任务。为了研究我们提出的框架在这项任务上的性能,需要在图像和诗歌匹配的地方构建一个图像-诗歌对的数据集。为了方便数据准备,我们重点生成四行七字的绝句。本文提出的框架可以很容易地推广到生成其他类型的诗歌。

为了构建图像-诗歌对的数据集,我们从互联网上收集了 68,715 幅图像,并使用了 (Zhang and Lapata 2014) 中提供的诗歌数据集,其中包含 65,559 个 7 字绝句。鉴于大量的图像和诗歌,手动匹配它们是不切实际的。相反,我们利用图像和诗歌的关键概念并自动匹配它们。具体来说,对于每张图像,我们使用 Clarifai 提供的 general-v1.3 模型获得几个关键词(例如,水、树);同样,我们从每一行诗中提取关键概念。然后可以匹配具有共同概念的图像和诗句。这样就构建了一个包含 2,311,359 个样本的数据集,每个样本由一个图像、前面的诗行和下一个诗行组成。我们随机选择 50,000 个样本进行验证,1,000 个样本用于测试,其余样本用于训练。

训练细节我们使用

6,000 个最常用的字符作为词汇表EG。编码器和解码器的循环隐藏单元的数量设置为 512。主题记忆网络中输入和输出记忆的维度也都设置为 512。平衡PT和PG的超参数 λ 设置为 0.5,在验证集上从 $\lambda = 0.0, 0.1, \dots, 1.0$ 。所有参数均从均匀分布随机初始化,支持 $[-0.08, 0.08]$ 。

该模型使用 AdaDelta 算法 (Zeiler 2012) 进行训练,批量大小设置为 128,并根据验证集上的交叉熵损失选择最终模型。在训练期间,我们按照 (Yi, Li, and Sun 2016) 反转要生成的每一行,以使模型更容易生成符合节奏规则的诗歌。对于视觉特征提取器,我们选择预训练的 VGG-19 (Simonyan and Zisserman 2014) 模型,并使用 conv5 4 层的输出,包括 512 维的 196 个向量,作为局部视觉特征图片。对于大多数图像,提取了 10-20 个关键字。

评估指标

对于自然语言处理中文本生成的一般任务,存在各种评估指标,包括 BLEU 和 ROUGE。然而,已经表明基于重叠的自动评估指标与人工评估几乎没有相关性 (Liu et al. 2016)。因此,对于自动评估,我们只计算生成的诗歌中描述的图像中关键概念的召回率,以评估我们的模型是否可以在给定图像的情况下生成一致的诗歌。考虑到独特性

表 2:所有模型的人工评估。粗体值表示最佳性能。

模型	诗意	流畅	连贯	意义	一致性	平均				
贴片机	6.97			5.85	5.15	6.62	5.08	6.20	5.41	5.52
RNNPG-A	7.27			7.69	6.70	6.78	5.15	4.81	3.98	5.91
RNNPG-H	7.36					5.51				5.67
PPG-R	6.47					4.91				4.82
PPG-H	6.52					5.24				5.09
MIPG (完整)	8.30					7.07				7.16
MIPG (无关键字)	7.21					6.13				6.15
MIPG (无视觉)	7.05					4.76				5.21

诗歌生成任务在文本结构和文学创作方面,我们评估生成诗歌的质量

与人类研究。

在 (Wang et al. 2016b) 之后,我们使用列出的指标
下面来评估生成的诗歌的质量:

- 诗意。这首诗是否遵循韵律和音调规定?
- 流畅。这首诗读起来流畅流畅吗?
- 连贯性。这首诗是否跨行连贯?
- 意义。这首诗是否有合理的含义和意境?

除了这些指标,对于我们的生成任务
中国古典诗词给定意象,我们需要评价
生成的诗歌如何传达输入图像。这里
我们引入了一个度量一致性来衡量是否
生成的诗歌的主题和给定的图像匹配。

型号变体

除了提出的 MIPG 框架外,我们还评估
模型的两个变体来检查

关于质量的视觉和语义主题信息
生成的诗歌:

- MIPG (完整)。所提出的模型,它集成了直接的视觉信息和语义主题信息。
- MIPG (无关键字)。基于 MIPG 的语义
通过将关键字的输入和输出内存向量设置为 ~0 来删除主题信息,这样模型
仅在诗歌生成中利用视觉信息。
- MIPG (无视觉)。基于MIPG,通过设置视觉特征向量来去除视觉信息
图像到~0,这样模型只利用语义
诗歌生成中的主题信息。

基线

据我们所知,之前没有从图像生成中国古典诗歌的工作。因此,对于基线,我们实
现了以下几种先前提出的基于关键字的方法:

贴片机。一种统计机器翻译模型 (He, Zhou,
和Jiang 2012),将前面的行转换为下一行,第一行是从输入生成的
使用基于模板的方法的关键字。

表 3:所有模型生成的诗歌中描述的图像中关键概念的召回率。

模型召回	模型召回		
RNNPG-A	12.85%	PPG-R	33.5%
RNNPG-H	11.82%	PPG-H	33.7%
SMT	19.79%	MIPG	58.8%

RNNPG。一种基于循环的诗歌生成模型
神经网络 (Zhang 和 Lapata 2014),其中第一个
一行是使用基于模板的方法生成的,以关键词为输入,其他三行是按顺序生成
的。在我们的实现中,使用了两种方案

选择第一行的关键字,包括使用所有
关键字 (RNNPG-A)并使用重要的关键字
分别由人类 (RNNPG-H)选择。具体来说,
对于 RNNPG-A,我们使用从
图片 (10-20 个关键词),而对于 RNNPG-H,我们邀请 3 个
志愿者投票选出最重要的关键词
图片 (3-6 个关键词)。
PPG。基于注意力的编码器-解码器框架

为每一行分配一个子主题关键字 (Wang et al.
2016b)。由于 PPG 要求关键字的数量
等于行数 (四行诗为 4 行),我们认为
从图像中提取的关键字集中选择4个关键字的两种方案,包括随机选择 (PPG-
R)
和人类选择 (PPG-H)。具体来说,对于 PPG-R,我们
从关键字集中随机选择 4 个关键字并随机排序。对于PPG-H,我们邀请3名志愿
者投票
对于图像中最重要的 4 个关键字并排序
它们按相关性顺序排列。

人工评价

我们邀请了十八名志愿者,他们在
中国古典诗歌从读到写,去评价
各种方法的结果。随机抽取 45 张图像作为我们的测试集。志愿者对每首生成的
诗歌进行评分
从 1 到 10 分,从 5 个方面进行评分:诗意性、流畅性、连贯性、意义和一致性。表
2 总结了结果。

整体表现。表 2 中的结果表明
所提出的模型 MIPG 优于基线
所有指标。值得一提的是,在衡量生成诗歌的好坏的“一致性”方面



 <p>扁舟一曲水平堤， I sing a fishing song on a boat in the lake overflowing its bank, 一棹渔舟日向西。 rowing oars with the sun setting in the west. 长忆西湖水中月， I often miss the moon reflected in the West Lake, 东风吹过武陵溪。 and the east breeze blowing across the WuLing River.</p>	 <p>春风庭院养花姿， Breeze blows beautiful flowers in the courtyard, 春入帘栊叶满枝。 Spring comes into my window, with leaves covering the branches. 堪笑门前青草树， Glad to see green grass and trees in front of my door, 谁家芳节几多时。 However spring will not last very long.</p>
---	---

表 4:由 MIPG 从相应图像生成的两首诗样本。

可以描述给定的图像,MIPG 实现了最好的性能,证明了所提出的图像的有效性
捕捉视觉信息和语义主题模型
生成的诗歌中的信息。从比较
在 RNNPG-A 和 RNNPG-H 中,我们可以注意到,通过手动选择重要的关键词,由
RNNPG-H 与图像更一致,这意味着
关键词选择在诗歌生成中的重要性。
同样,在手动选择重要关键字的情况下,PPG H 在各个方面都优于 PPG-R,尤其是在“Coherence”和“Consistency”方面。作为比较,在提出的 MIPG 框架中,主题记忆网络使
可以在生成每个字符的同时动态确定一个主题,同时,编码器-解码器
框架确保按照规定生成流利的诗歌。作为一个整体,视觉信息和
主题记忆网络共同生成诗歌
与给定的图像一致。

模型变体分析。结果在底部
表 2 的行对应于模型变体,包括 MIPG (完整)、MIPG (无关键字)和 MIPG (无视觉)。可以观察到忽略语义关键字
MIPG (无关键字)或 MIPG 中的视觉信息 (无视觉)显着降低了所提出模型的性能,尤其是在“一致性”方面。这提供了明确的证据证明视觉信息
和语义主题信息共同生成
诗与形象一致。

意象-诗歌一致性的自动评价

不同于基于关键词的诗歌生成模型,
对于从图像生成诗歌的任务,图像诗歌一致性是一个新的非常重要的指标,而
评估生成的诗歌的质量。因此,我们
根据“一致性”进行自动评估
除了人工评估,这是由 com 实现的

将关键概念的召回率放在图像中
在生成的诗中描述。
结果如表 3 所示,其中可以观察到所提出的 MIPG 框架优于所有
余量较大的其他基线,这表明
MIPG 生成的诗歌可以更好地描述
给定的图像。人们还应该注意两者之间的区别
“一致性”中的主观和定量评价
来自表 2 和表 3。例如,PPG-R 和 PPG-H
实现了几乎相等的关键词召回率,但 PPG-R 的“一致性”得分明显
低于 PPG-H
在表 2 中。考虑到 PPG-R
PPG-H 选择从给定图像中提取的 4 个关键字来生成诗歌,产生相同的召回率。在里面
同时,人工选择的关键字比随机选择的关键字更能描述图像,因此 PPG-H
从人类的角度来看,可以生成比 PPG R 更符合图像的诗歌。此外,原因在于
RNNPG-A 和 RNNPG-H 的关键词召回率为
相对较低可能是由于 RNNPG 经常
生成第一首语义相关的诗行
使用给定的关键字,但不包含任何关键字。作为一个
因此,RNNPG-A 和 RNNPG-H 可能会生成关键词召回率低但“一致性”高的诗歌
分数。

例子

为了进一步说明由
提出的 MIPG 框架,我们包括两个例子
用表 4 中的相应图像生成的诗歌 如这些示例所示,由 MIPG 生成的
诗歌
可以捕捉视觉信息和语义概念
给定的图像。更重要的是,这些诗以诗意的方式很好地描述了图像,同时遵循
中国古典诗歌的严格规定。

结论在本文中,我

们提出了一种基于记忆的神经网络模型,用于从图像生成中国古典诗歌 (MIPG),其中在生成诗歌时利用图像的视觉特征和语义主题。给定一幅图像,语义关键词被提取为诗歌的骨架,其中提出了一个主题记忆网络,该网络可以获取尽可能多的关键词,并在诗歌生成过程中动态选择最相关的关键词。在此基础上,整合视觉特征,体现关键词中缺失的信息。从不同角度对诗歌质量的数值和人类评估证明,所提出的模型可以生成以诗意的方式准确描述给定图像的诗歌,同时遵循中国古典诗歌的严格规定。

致谢本研究得到国家自然科学基金

基金 (No. 61375060、No. 61673364、No. 61727809 和 No. 61325010)和中央高校基本科研业务费专项资金 (WK2150110008)的资助。

我们也非常感谢 NVIDIA 公司捐赠用于这项工作的 Titan X GPU 的支持。

参考文献Cho,

K.;范梅林博尔,B.;古尔切赫尔,C.;巴赫达瑙,D.布加雷斯,F.施文克,H.和 Bengio, Y. 2014。使用 rnn 编码器-解码器学习短语表示以进行统计机器翻译。 arXiv 预印本 arXiv:1406.1078。

他,J.周,M. and Jiang, L. 2012。使用统计机器翻译模型生成中国古典诗歌。在 AAAI。

Jiang, L. 和 Zhou, M. 2008。使用统计 mt 方法生成中国对联。在第 22 届计算语言学国际会议论文集第 1 卷,377-384。计算语言学协会。

刘,C.-W.降低。;塞尔班四世;值得一提的是,M.查林,L.;和 Pineau, J. 2016。如何不评估您的对话系统:对话响应生成的无监督评估指标的实证研究。 arXiv 预印本 arXiv:1603.08023。

曼努隆,R.;里奇,G.和 Thompson, H. 2012。使用遗传算法创建有意义的诗歌文本。实验与理论人工智能杂志 24(1):43-64。

Manurung, H. 2004。诗歌生成的进化算法方法。

内策尔,Y.加贝,D.戈德堡,Y.和 Elhadad,M. 2009。Gaiku:使用单词关联规范生成俳句。在关于语言创造力的计算方法研讨会的会议记录中,32-39。计算语言学协会。

Oliveira, H. 2009。诗歌的自动生成:概述。科英布拉大学。

Oliveira, HG 2012。Poetryme:诗歌生成的多功能平台。计算创造力、发明概念和通用智能 1:21。

Simonyan, K. 和 Zisserman, A. 2014。用于大规模图像识别的非常深的卷积网络。 arXiv 预印本 arXiv:1409.1556。

苏赫巴托尔,S.;韦斯顿,J.弗格斯,R.等。 2015。端到端内存网络。在神经信息处理系统的进展中,2440-2448。

;土佐,N.小原,H.和 Minoh, M. 2008。Hitch haiku:用于创作俳句的交互式支持系统。在娱乐计算国际会议上,209-216。施普林格。

王,Q.罗,T.王,D.和 Xing, C. 2016a。基于神经注意力模型的中文歌曲抑扬格生成。 arXiv 预印本 arXiv:1604.06274。

王,Z.他,W.吴,H.吴,H.李,W.王,H.和陈,E. 2016b。基于规划的神经网络的中文诗歌生成。 arXiv 预印本 arXiv:1610.09889。

Wang, L. 2002。中国诗歌押韵约束概述。

韦斯顿,J.乔普拉,S.和 Bordes, A. 2014。记忆网络。 arXiv 预印本 arXiv:1410.3916。

吴,X. ;土佐,N.和 Nakatsu, R. 2009 年。新搭扣俳句:应用于观光导航系统的交互式 renku 诗歌创作支持工具。在娱乐计算国际会议上,191-196。

施普林格。

严,R.江,H.拉帕塔,M.林,S.-D.吕,X.和 Li, X. 2013。i, 诗人:通过约束优化下的生成摘要框架自动进行中国诗歌创作。在 IJCAI。

易,X.李,R.和 Sun, M. 2016。使用 rnn 编码器-解码器生成中国古典诗歌。 arXiv 预印本 arXiv:1604.01537。

Zeiler, MD 2012。Adadelata:一种自适应学习率方法。 arXiv 预印本 arXiv:1212.5701。

Zhang, X. 和 Lapata, M. 2014。使用循环神经网络生成中国诗歌。在 EMNLP,670-680。

张,J.冯,Y.王,D.王,Y.亚伯,A.张,S.和 Zhang, A. 2017。使用神经记忆的灵活和创造性的中国诗歌生成。 arXiv 预印本 arXiv:1705.03773。

周,C.-L.你,W. and Ding, X. 2010。中国宋词自动生成的遗传算法及其实现。软件杂志 21(3):427-437。