

基于双向多层转换编解码的诗自动生成

杨泰康, 杨婉霞, 刘燕, 胡智喻, 王巧珍, 徐明杰

(甘肃农业大学机电工程学院, 甘肃 兰州 730070)

✉2605946368@qq.com; yangwanxia@163.com; 751741004@qq.com;
hzybest@foxmail.com; 3146477442@qq.com; 1004106102@qq.com



摘要: 中国古诗因其严谨的平仄、押韵结构, 精练传神的用词成为文本自动生成领域的挑战问题之一。本文基于双向多层转换编解码的空间向量, 结合Attention机制建立了循环神经网络结构的诗自动生成模型。为解决自动生成诗的主题发散性问题, 模型在生成每句诗时均增加了关键词约束。此外, 为了增强诗句之间的对仗性和语义的连贯性, 模型双向多层转换编解码的词嵌入式输入添加了诗句的对齐向量。实验结果表明, 相比于以词转换为向量的词嵌入式的诗自动生成模型, 本文设计的基于BERT的对齐向量输入模型生成的诗不仅在机器评估和人工评估中性能较优, 而且生成诗句相邻句子的相关性也最好。这也进一步说明, 当模型的输入向量能充分表达诗词的格式、内容和语义时, 双向多层转换编解码的向量表示, 即Attention机制+循环神经网络构建的诗生成模型可以生成较接近人工作的诗。

关键词: Attention机制; 双向多层转换编解码; 诗; 自动生成; 循环神经网络

中图分类号: TP391 文献标识码: A

Automatic Generation of Poem based on Two-way Multi-layer Conversion Codec

YANG Taikang, YANG Wanxia, LIU Yan, HU Zhiyu, WANG Qiaozhen, XU Mingjie

(School of Mechanical and Electrical Engineering, Gansu Agricultural University, Lanzhou 730070, China)

✉2605946368@qq.com; yangwanxia@163.com; 751741004@qq.com;
hzybest@foxmail.com; 3146477442@qq.com; 1004106102@qq.com

Abstract: Chinese ancient poems have become one of the challenges in the field of automatic text generation because of their tonal patterns, rhyming structure, and vivid words. This paper proposes to establish an automatic poem generation model with a recurrent neural network structure. This model is combined with Attention and is based on space vector of two-way multi-layer conversion codec. In order to solve the problem of divergence of automatically generated poems, the model adds keyword constraints when generating each poem. In addition, in order to enhance antithesis between verses and semantic coherence, word embedded input of the model's two-way multi-layer conversion encoding and decoding adds alignment vector of the poem. The experimental results show that compared with word-embedded poem automatic generation model in which words are converted into vectors, poems generated by BERT-based aligned vector input model proposed in this paper, not only perform better in machine and manual evaluation, but also generate poetic sentences. The correlation between adjacent sentences and the generated verse is also the best. This also further shows that when input vector of the model can fully express the format, content and semantics of the poem, vector representation of the two-way multi-layer conversion codec, the Attention Mechanism + cyclic neural network construction of poem generation model can generate poems closer to human's works.

Keywords: Attention Mechanism; two-way multi-layer conversion codec; poem; automatic generation; Recurrent Neural Network

1 引言(Introduction)

中国古诗词源远流长, 大约有2000多年的历史, 是中国

古典文学中最璀璨的明珠。不同时期的中国古诗词主要发展为唐诗、宋词和元曲等不同类型。这些古典诗词中, 以格律

诗为主，主要是五言和七言的绝句和律诗，其突出特点是结构严谨，字数、行数、平仄、用韵都有一定的限制。例如，律诗一般讲究平仄和押韵、押韵和对仗，其中最核心的是它的韵律，即所谓的押韵，就是在诗词中的规定位置(一般在句末)放置相同韵部的字，以使声韵和谐。句子押韵，不仅吟诵和记忆舒畅，更使作品具有节奏、声调和谐之美，如“平平仄仄平平仄，仄仄平平仄仄平”。杜甫的《春望》是一首标准的律诗：

国破山河在，(仄仄平平仄)
城春草木深。(平平仄仄平)
感时花溅泪，(平平仄仄)
恨别鸟惊心。(仄仄仄平平)

其中第二句的“深”和第四句的“心”押韵，第三句和第四句对仗。中国古诗词除了传承经典的文化精髓，还有教育和启示后人的作用。而在当今信息和数据时代，诗词也被赋予了新的使命，它不仅是构成信息和数据的一员，更是承载信息的主要载体之一。所以机器创作和自动生成诗词成为研究的热点，它不但可以传承和发展古代文化，更能满足普通人进行诗词创作的意愿，进一步激发人们学习和创作诗词的热情。然而由于受中国传统诗词严格的平仄押韵等格律限制，诗词自动生成极具挑战性。但随着计算机技术和机器学习的发展，使诗词的创作方式和生成手段发生了前所未有的改变，特别是深度学习的再次崛起，使诗词创作成为可能，这也是本文关注的研究方向所在。

自20世纪90年代中期开始研究计算机辅助创作古诗词以来，已在语料库建立^[1-2]、词汇语义^[3]、文本自动生成^[4]等方面取得了一定的研究成果，并且能用多种方法生成符合一定格式要求的诗词。但是，生成的诗词不具有流畅性和连贯性，更没有感情色彩。其主要原因是传统的诗词生成模型对生成诗词的格律、语义等的约束一般由人工设计的规则和评估函数来实现，很难全面地兼顾词与词、句子与句子、主题与内容之间的相关性。近年来，随着深度学习技术在自然语言处理方面的深入研究，出现了基于卷积神经网络(Convolutional Neural Networks, CNN)和循环神经网络(Recurrent Neural Network, RNN)的诗词生成模型，大大地提高了诗词的生成质量。由于RNN存在梯度消失问题，对长序列的学习效果不佳，改进的门控单元(Gate Recurrent Unit, GRU)虽然能克服梯度消失的问题，然而由于不论输入长短，它都将其编码成一个固定长度的向量表示，这样使得模型对长输入序列的学习效果依然较差。后有研究者尝试利用语义图作为深度神经网络的输入以提高生成文本的主题一致性和可控性，但还只是初步研究阶段，有待进一步深入研究。所以本文提出一种基于BERT(Bidirectional Encoder Representation from Transformers，双向多层转换编解码)词向量的BiGRU+

Attention机制的编解码模型来自动生成诗词。实验结果表明，本文设计的BERT词向量结构模型能够生成具有流畅性、连贯性的诗歌。

论文其余部分结构为：第二部分介绍了诗词自动生成的研究现状。第三部分构建了本文诗词生成的结构过程并解释了各模块的功能实现原理。第四部分详细描述了模型的训练过程。第五部分是实验的设计和实验结果的评估。第六部分是结论。

2 相关研究(Related research)

纵观诗词自动生成发展过程，可将诗词的生成大致分为三个阶段：第一阶段是基于模板和模式的生成方法，该方法简单可靠，能输出高质量的结果，如利用词的关联规则生成日语诗^[5]，应用语义和语法模板生成西班牙语诗^[6]。但这种方法不够灵活，需要人工设计模板或模式，虽然后来的基于实例推理方法的模板是从现有诗词库检索生成，但过程依然繁琐。第二阶段的诗词生成主要以自然选择为主，即利用遗传进化算法优化生成，如采用随机搜索算法获得更具匹配性诗句的方法^[7]，但这类算法需要精心设计评估函数来保证诗词的语义性、连贯性和格律等要求，不仅难度大，生成诗词的效率也较低。第三阶段的诗词生成主要基于统计机器翻译(Statistical Machine Translation, SMT)算法，SMT将诗词生成看成翻译过程，首次成功应用于对联的生成^[8-9]。该类生成方法主要的遗憾是无法生成诗的第一句。上述方法在不断地提高和完善的过程中，基本实现了符合一定语义性、连贯性和格律要求的诗词自动生成系统，但缺乏通用性，迁移能力比较差。随着深度学习技术的进一步发展，通过训练深层次的网络对大量的诗词进行学习，将学习到的诗词的语义信息和格律训练成模型，然后利用该模型生成具有一定主题的诗词，去除了人工的参与和设计，真正意义上实现了诗词的自动生成。如基于RNN神经网络的诗词生成模型^[10]能够自动学习诗词的字词表示、组合选择及诗句的内容。然而它的结构有些复杂，由一个CNN和两个RNN构成，不利于模型的扩展。有文献提出了Attention机制的机器翻译模式的宋词自动生成方法^[11-12]，Attention机制的引入增强了语义的连贯性，但存在主题漂移的问题。为此，基于主题规划的诗歌生成模型^[13]解决了主题发散的问题，但缺乏自动评估。相关文献将对抗式生成网络(Generative Adversarial Nets, GAN)引入文本生成中^[14-15]，提出了具有生成器RNN和判别器CNN的SeqGAN(Sequence Generative Adversarial Nets)诗歌生成模型，解决了评估函数难以人工设计的问题，但还有很多值得探索和改进的地方。IOANNIS K等^[16]直接将语义图(Abstract Meaning Representation Graphs, AMR Graphs)作为seq2seq模型的输入生成文本。它首先是利用深度优先搜索将AMR变成序列，对该序列采用LSTM模型进行编码，然

后基于Encoder–Decoder生成文本。这种方法并没有直接对图形结构建模，而是对图形结构的线性和序列化编码。SONG等^[17]基于Graph LSTM(Long Short-Term Memory，长短期记忆)的门控图神经网络(Gated Graph Neural Network，GGNN)对输入图形结构进行直接编码来生成文本。但门控网络会影响信息传播，生成的文本主题出现不确定性。后GraphWriter模型对图注意力网络进行了扩展^[18]，能够同时对文章标题和提取的知识图进行编码，解码时输出的概率表示与知识图和标题相互关联，使生成文本在主题一致性上有所改进，但生成文本的上下文语义连贯性和主题可控性方面均有待提高。由此可见，高质量诗词的自动生成仍然需要更进一步的研究。由上述分析可知，目前还没有基于BERT词向量的诗词生成类的相关报道。为此，本文提出了BERT嵌入式词向量的、关键词监督的、Attention机制的对齐模式诗生成模型。其主要特点是BERT在将文本转换为空间向量时表达了动态的句子级语义信息，计算输出的是实时的特征矢量。第二，每句诗的生成都有相应的关键词引导，不仅仅受第一句的影响，以保持主题的一致性。第三，添加对齐序列向量输入以增加生成诗句之间的对仗性，使生成的诗句保持语义上的连贯性和结构的流畅性。本文的主要贡献点是将BERT动态词向量引入诗生成模型，解决了模型对输入句子的句子级语义理解问题。

3 诗词生成(Generation of poems)

考虑到生成诗词的主题一致性问题，本文诗词生成主要分为三个模块，即关键词提取与扩展模块(Keyword–Extract–Extend， KEE)、字生成诗句模块(Word–To–Line， WTL)和关键词引导的Attention机制诗句生成模块(Context–To–Line， CTL)。诗词的整个生成过程如图1所示。下面将具体介绍每个模块的构建和功能。

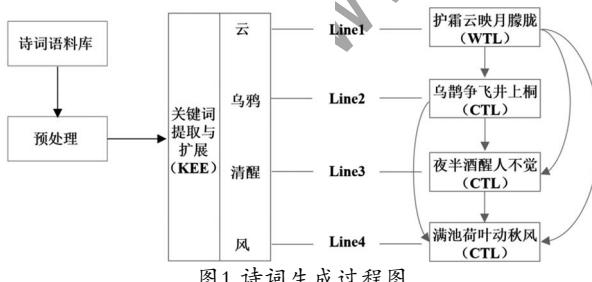


图1 诗词生成过程图

Fig.1 Graph of poem generation process

3.1 关键词提取与扩展模块(KEE)

本文诗句生成模型的训练语料是关键词加诗句结构，那么形成训练语料的关键一步是提取每句诗的关键词。目前，文本关键词的提取分为无监督和有监督方法，有监督方法需要有诗词语料的关键词标注数据的支持，但由于人工标注的成本较高，因此本文采用无监督的关键词提取方法。其过程是：首先对收集的诗词语料采用结巴和诗学含英进行分词，

这里结合诗学含英分词，主要是考虑到诗句中字词语义表达的特殊性。然后用TextRank算法对诗句中的候选词打分并排序，之后选取得分较高的N个词作为关键词，同时保留这些词在原句中的顺序。

关键词的扩展是指当用户输入的关键词太少，或已知句子太短无法提取出足够的关键词时，可采用输入的关键词或已提取的关键词训练Word2vec(Word to Vector，词转换为向量)预训练模型生成词向量，然后利用词向量计算相似度，获得语义最近的词，将其扩展为关键词。作诗时，用户根据要表达的主题和思想选定关键词并输入模型，此时，关键词扩展模块会根据用户输入的关键词将其扩展到N(假设有N个诗句)。或者用户可以输入一个句子，关键词提取模块可以从用户输入的句子中提取关键词并扩展至N个。因为本文此次训练模型生成的是四言诗，所以最多扩展至四个关键词。

3.2 字生成诗句模块(WTL)

由研究现状分析可知，基于统计(SMT–based)的机器翻译方法是生成诗歌综合性能较高的方法之一，但其最大的不足是自身无法生成第一句诗，其主要原因是语言要素的合理表达方式的欠缺。而本文模型结构采用BERT将所有输入转化为嵌入式词矢量，那么词和句子都将映射到相同的矢量空间，这些空间中的矢量表达了词向量、词的位置向量和分段向量的和，具有更深和更广的语义关系。这样，就可以将诗词的生成系统简化。也就是说，理论上本文的字生成诗句和诗句生成诗句这两个模块都可以基于Attention机制的RNN Encoder–Decoder一个结构实现，但考虑到首句的生成只是依据关键词，输入相对较少，所以将该WTL模块单独训练。其具体过程是根据扩展后的关键词，WTL模块生成与输入的关键词相关语义的首句。之后CTL模块基于生成的第一句和第二句的关键词生成诗的第二句。最后CTL以之前所有的生成诗句和本句关键词生成整个诗词。

3.3 关键词引导的Attention机制诗句生成模块(CTL)

基于Attention的seq2seq模型是由文献[19]首次应用到自然语言处理中(NLP)，之后出现了各种基于RNN的Attention的扩展模型，其主要是对RNN单元的改进和Attention矩阵计算方法变换的两类扩展模式。不管哪种扩展模式，它们的基本结构都是相似的。本文以关键词为导向的Attention机制诗句生成模块结构如图2所示。其中，BERT作为模型的词嵌入用来提取诗句中词的动态语义信息。它使用的是Transformer，捕捉到的是真正意义上的Bidirectional Context信息。Encoder部分由双向GRU对输入的诗词向量进行编码，Decoder的解码由GRU+Attention机制完成，Attention的主要贡献点在于根据与输出的匹配程度，对GRU隐藏层的状态进行加权变化，区别对待。本文模型的输入数据除了上下文信息，特别添加了关键词，即图2中的 $a_{t,0}$ ，以增

强主题的信息量。

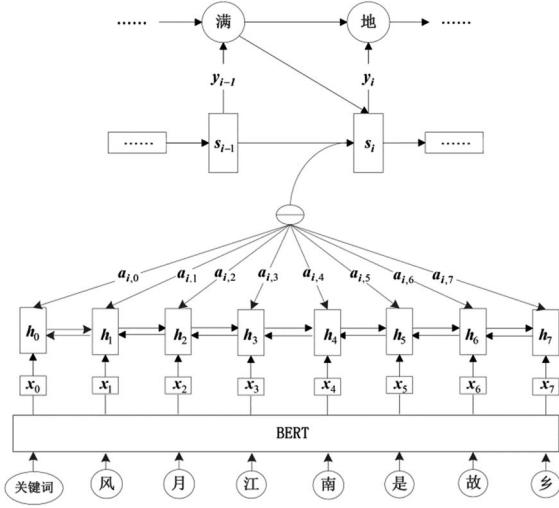


图2 BERT结合Attention机制的生成诗句结构图

Fig.2 The structure diagram of the poem generated by the Attention Mechanism combined with BERT

该模块的具体工作过程为：BERT将输入的文本信息表示为深层语义向量 (x_1, x_2, \dots, x_i) ，Encoder将输入序列 (x_1, x_2, \dots, x_i) 转换为正反向隐藏状态 $\bar{h}_j = f(x_j, \bar{h}_{j-1})$ 和 $\bar{h}_j = f(x_j, \bar{h}_{j+1})$ ，对于任意的 x_i ，通过编码形成的最终隐藏状态是相应的正反向隐藏状态的结合 $h_j = [\bar{h}_j, \bar{h}_j]^T$ ，这里， f 是非线性函数， $h_j \in R^{2n}$ 。这样，隐藏状态同时涵盖了词以及诗句的前后向语义关系，将获得的词语和诗句的全面完整信息用于后端的解码。在本文模型中，诗的第一句仅由关键词生成，即 $a_{i,0} = h_{j,0}$ ，由WTL模块完成。Decoder的解码端根据自身当前的状态 s_i 和编码端输出的隐藏状态 h_j ，以字符的形式生成诗句 (y_1, y_2, \dots, y_i) 。两个模块(WTL、CTL)的解码端功能相同。具体公式为：

$$y_i = W_o S_i + b_o \quad (1)$$

$$S_i = g(y_{i-1}, S_{i-1}, c_i) \quad (2)$$

其中， c_i 是注意力机制中用来计算第 i 个字符的隐藏状态 S_i 的，即

$$\begin{aligned} a_{i,j} &= \frac{\exp(v_{i,j})}{\sum_{j=1}^T \exp(v_{i,j})} \\ c_i &= \sum_{j=1}^T a_{i,j} h_j \\ v_{i,j} &= v_a^T \tanh(W_a S_{i-1} + U_a h_j) \end{aligned} \quad (3)$$

这里的 $a_{i,j}$ 反映了编码端隐藏状态对生成诗句的贡献大小，表达了 s_i 和 h_j 的相似度，也称为对齐度。解码器依据 $v_{i,j}$ 确定哪一部分输入更加重要，以分配更多的注意力，从而使生成诗的上下句之间的相应词具有更好的关联性。

4 模型的训练(Model training)

4.1 词嵌入向量模型的训练

词嵌入层是诗生成模型的输入层，也是诗生成过程中最关键的一步。该层能否将词或句子的语义和感情等特征映射到相应的语义空间，将直接决定整个生成模型的性能。考虑

到诗词的稀疏性以及诗句结构的特殊性(如对仗、押韵等)，本文的词嵌入向量采用BERT生成，如图3所示。

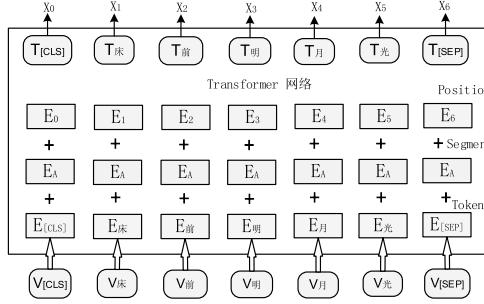


图3 BERT结构模型图

Fig.3 The structure model diagram of BERT

通过训练，BERT提取出输入序列 (v_1, v_2, \dots, v_i) 的token、位置和分割嵌入特征，将它们的特征和转换为具有深层语义关系的字向量 x 输出。本文基于BERT训练的模型主要有两种，分别为对齐模型和普通模型。其中对齐输入的词向量模型是本文的一个突出点，主要是为了提取和学习诗句间的对仗性，即将诗句结合其对应的对仗数据列表共同作为BERT词嵌入向量模型的训练数据，以得到诗上下句间更丰富的语义信息。具体为：第一个列表存放这首诗每一句的第一个字，第二个列表存放该诗每一句的第二个字，以此类推直到所有字放入列表中，列表的个数由诗的体裁决定，如表1所示。而普通模型的词向量仅用诗句训练BERT模型，将其转化为嵌入式词向量。

表1 对仗数据列表生成

Tab.1 Generation of antithesis list

原诗	对仗数据列表
	['故', '双', '马', '凭']
故园东望路漫漫，	['园', '袖', '上', '君']
双袖龙钟泪不干。	['东', '龙', '相', '传']
马上相逢无纸笔，	['望', '钟', '逢', '语']
凭君传语报平安。	['路', '泪', '无', '报']
	['漫', '不', '纸', '平']
	['漫', '干', '笔', '安']

4.2 混合训练

就句型而言，古诗一般有五言和七言之分，即每句诗有五个或七个字符。虽然在格式上它们有所不同，但意境、思想的表达所采用的字词是相似的。为了丰富训练集，参考已有文献，选择采用五言和七言的混合诗集来训练同一模型。

4.3 诗生成模型的训练

诗生成模型的训练就是利用训练集，通过最优化方法确定一组使损失函数取值最小的参数，确定参数后的函数就是训练的结果。使损失函数取值最小的直接表达就是使预测值和原始值更加接近。就诗词生成而言，是指生成诗句符合训练诗句的格式，输入句和生成句的主题保持一致，内容关联，富含韵律和对仗。为此，本文选择交叉熵损失函数来衡量Decoder输出字符和真实输入值之间的误差。另外，为了提高模型训练的效率，选择最小批梯度下降(Mini-batch

Gradient Descent)算法加快迭代次数和AdaDelta算法优化学习效率。通过多次试验，主要参数确定为batchsize为64、dropout为0.3、学习率为0.0001的训练效果最优。

5 实验及方法评估(Experiment and method evaluation)

这部分主要是数据集的获取、实验方法的设计和实验结果的评估。首先介绍数据集的构成，其次是实验方法以及与其他方法的比较，最后是对实施方法的评估。

5.1 数据集

本文的数据集是从互联网上搜集的五言和七言律诗，其中五言四句诗有96,208首，七言四句诗有175,603首，共271,811首，用来训练GRU+Attention模型。据我们所知，该数据集已经涵盖了大量已有的诗句，并从中随机选择4,000首诗作为验证集、4,000首诗作为测试集，剩下的诗为训练集。

5.2 实验步骤

本文诗生成实验的具体步骤如下：

(1)首先对收集的数据集进行预处理，去除停用词，按照词频排序，生成词典。

(2)对语料库中的诗进行解析，判断其是否为五言或七言律诗以及这些诗是否在生成的词典中。

(3)对解析后的所有诗按照3.1节的方法提取并扩展关键词，对于每一行诗句，使用TextRank的方法选出每行得分最高的词作为该行的关键词。本文的训练集为绝句，所以每首绝句被提取出有序的四个关键词用来训练关键词扩展模型。

(4)用预处理的诗语料对BERT词嵌入模型进行训练，得到训练语料的词向量空间。

(5)训练集的生成。按照图1的诗生成方式处理并生成训练集，如表2所示。

(6)采用4.3的参数及生成的训练集训练BERT+BiGRU+Attention序列模型。

表2 训练集的生成

Tab.2 Generation of training set

诗	关键词	上文
护霜云映月朦胧	月	
乌鹊争飞井上桐	乌鹊	护霜云映月朦胧
夜半酒醒人不觉	酒醒	护霜云映月朦胧；乌鹊争飞井上桐
满池荷叶动秋风	风	护霜云映月朦胧；乌鹊争飞井上桐；夜半酒醒人不觉

5.3 实验设置

实验重点设计了基于BERT对齐向量输入的Attention机制诗生成模型，简称BERT对齐模型(B_AM)。为了对比BERT对诗及句子语义的动态深层次表示，实验还训练了基于Word2vec的对齐向量输入模型(W_AM)，以及目标句反转输入模型，简称反转模型(W_RM)和普通模型(W_CM)。其中反转模型是指在训练时将目标句反转输入，其主要目的是为

了增强上下诗句之间的局部一致性，但其负面影响是削弱了对诗句的韵律表征。普通模型是指按照诗句的正常顺序输入Word2vec及序列模型进行训练。对每种模型，实验均采用五言和七言诗混合训练集进行训练，本文系统是在Disiok的基础上建立的。同时，为了对比本实验设计的神经网络诗生成模型的优劣性，将此次实验设计的四种模型(B_AM、W_AM、W_RM、W_CM)与SMT以及无Attention机制的模型(No-Attention，NAT)在同一主题下作的诗进行了对比，其结果主要通过BLEU(Bilingual Evaluation Understudy，双语互译质量评估辅助工具)和人工两方面进行评测。

5.4 诗的生成及评估

实验按照诗学含英中对诗词的分类，分别选取了思乡、边塞、抒情、叙事和山水田园五类主题的诗各5首，从每首诗中人工选取1—4个关键词，分别用实验设置中的模型及对比模型作诗，每类主题每个模型作10首诗，每个模型共作各类主题的诗50首，人工初选出各模型的30首作评测。此次采用的评估方法主要有以下三种。

5.4.1 BLEU评测

首先对各模型生成的30首诗采用BLEU进行评测。BLEU原本是用来评价机器翻译效果的，也有用BLEU评测诗词的生成质量。该方法简单，计算量小，最主要的是与人工评测的结果有一定的相似性，所以本文首先采用BLEU对生成的诗进行自动评估。表3是本文四种模型与SMT和NAT模型生成诗的BLEU结果。可以看出，BERT对齐模型生成诗的平均BLEU最高，其次是对齐向量输入模型。而SMT模型生成的诗的BLEU相对较低，因为其主要采用模板或统计模型的填词形式生成诗，所以生成诗句的语义连贯性不强，可读性较低。

表3 各模型的BLEU值

Tab.3 BLEU value of each model

Model	B_AM	W_AM	W_RM	W_CM	SMT	NAT
BLEU	0.551	0.544	0.445	0.443	0.212	0.400

5.4.2 相关性评测

对于自动生成的诗，除了格式上符合要求以外，更重要的是语义上的流畅和连贯性。目前，这种性能的机器评测大多采用上下句的相关性进行衡量。本文借鉴BERTScore对自动生成文本或摘要的评价思想，即对两个生成句和参考句分别用BERT提取特征，然后对这两个句子的每一个词分别计算内积，可以得到一个相似性矩阵。基于这个矩阵，计算参考句和生成句的最大相似性得分的累加然后归一化。文本在此思想上做了改进，侧重于比较诗的上下句子之间的相似度，以衡量生成整首诗的流畅性。根据前面BLEU的评测结果，择优比较，本实验只选取B_AM和W_AM两种对齐模型和人工作的同一主题诗的上下句的相关性做了对比分析，结果如表4所示。为了更直观地表达三者的关系，作了如图4所示的曲线图，从表4和图4可以看出，基于BERT的对齐模型的诗句间的

相关性要好于Word2vec对齐模型生成诗句的相关性，仅次于人工作诗。

表4 基于BERT词嵌入和Word2vec词嵌入与人工诗句的相似度值

Tab.4 Similarity value of manual verses and verses based on BERT word embedding and Word2vec word embedding

主题	句子	B_AM	均值	W_AM	均值	人工	均值
思乡	1—2	0.266		0.251		0.274	
	2—3	0.257	0.257	0.241	0.245	0.275	0.266
	3—4	0.248		0.243		0.250	
边塞	1—2	0.254		0.235		0.269	
	2—3	0.246	0.253	0.234	0.239	0.232	0.245
	3—4	0.259		0.248		0.235	
叙事	1—2	0.258		0.223		0.259	
	2—3	0.243	0.25	0.236	0.227	0.259	0.261
	3—4	0.249		0.221		0.265	
抒情	1—2	0.263		0.232		0.262	
	2—3	0.251	0.258	0.242	0.236	0.262	0.266
	3—4	0.260		0.235		0.275	
山水田园	1—2	0.266		0.226		0.250	
	2—3	0.247	0.253	0.243	0.238	0.226	0.266
	3—4	0.246		0.246		0.283	

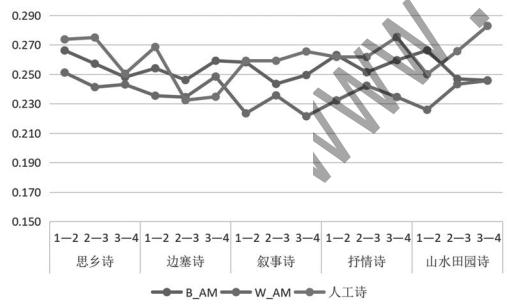


图4 相似度比较
Fig.4 Similarity comparison

5.4.3 人工评测

众所周知，文本生成系统还没有权威的、非常准确的评估标准，因为在已知上句时，有很多符合逻辑及语义关系的下句可以对应。而且，诗除了具有一般文本的相关属性外，还应有其特殊的严谨的结构，如平仄、押韵等，更重要的是诗的用意，这一点是无法用BLEU评测的。目前最好的办法是人工评测，本文参考之前的相关工作(文献[10]和[11])制定了五条人工评测标准：诗意图性、流畅性、连贯性、意义性及声调和押韵的和谐性。因为诗区别于一般文本的主要特点是它的结构和声调及押韵等，所以本文在人工评测中加了声调和押

韵的和谐性。其中，每一条标准的打分范围是3—5分。将每种模型生成的30首诗由五名文学专家分别对每条标准进行打分，计算它们的平均得分作为最终结果，如表5所示。

表5 各模型生成诗的人工评测结果

Tab.5 Manual evaluation results of poems generated by different models

标准	B_AM	W_AM	W_RM	W_CM	NAT	SMT
诗意图性	4.01	3.98	3.95	3.87	3.81	4.02
流畅性	3.96	4.02	3.94	3.88	3.83	3.98
连贯性	4.05	4.02	3.95	3.86	3.81	3.99
意义性	4.07	4.08	4.06	3.98	3.88	4.06
和谐性	3.92	3.89	3.81	3.67	3.63	3.85

从表5可以看出，本文的BERT词嵌入式对齐模型在后四个指标中较SMT有一定的优势，在诗意图方面有些欠缺，可能的原因是SMT的每一句诗的词是择优选择填充，甚至SMT的第一句可能是原诗而并非机器学习自动生成。但较Word2vec对齐模型、反转模型、普通模型和NAT模型，BERT对齐模型的优势较为明显，NAT模型整体性能欠佳。

6 结论(Conclusion)

本文采用sequence-to-sequence结构生成诗句序列，基于GRU+Attention机制建立诗的生成模型。为了从句子级别表达上下文动态的语义信息，本文自动生成模型的向量转换采用BERT词嵌入。为解决自动生成诗的主题发散性问题，建立了关键词导向的Attention机制的诗生成模型，即在生成每一句诗时均有关键词的监督和引导。为了增强诗句之间的对仗性和语义的连贯性，在原有数据(诗句)的基础上增加了对仗数据列表作为输入训练BERT模型生成嵌入式词向量，将该词向量作为诗生成模型的输入生成诗的下句。为了证明本实验设计模型生成诗的优劣性，将此次实验设计的三种模型(B_AM、W_AM和W_RM)与经典SMT模型在同一主题下作的诗进行了对比，主要从自动(BLEU)、人工和上下诗句的相关性三方面进行评估。从对各模型所作诗的各种评估结果分析可知，无论是BLEU值与原诗的对比评估，还是人工的五方面指标的评估，BERT对齐模型生成诗的效果均较优。特别是对生成诗的上下句相关性而言，BERT对齐模型优势明显，仅次于人工作诗；其次是Word2vec对齐模型；Word2vec反转模型由于在音调上的控制，整体性能高于普通模式。这也进一步说明，只要能使模型的输入序列充分表达诗句的格律、押韵和对仗，以及丰富的上下文内容和语义等信息，Attention机制可以生成较接近人工的诗。另外，从对比模型(NAT)的各方评测可以看出，基于神经网络的深度学习模型生成文本的质量关键在于语义的转换和编码部分，如果模型的编码输入部分能获得全面的、深层次的语义信息，结合Attention机制，深度学习模型可自动生成主题一致、满足格律要求的诗等各类文本，那么未来的工作可以尝试BERT预训练的文本或诗生成模型的建立及实验。

参考文献(References)

- [1] 林倩,文华婷,杨静,等.中文词语内部层次结构标注语料库的建立[J/OL].厦门大学学报(自然科学版),2020,59(02):225–230.
- [2] 刘扬,林子,康司辰.汉语的语素概念提取与语义构词分析[J].中文信息学报,2018,32(02):12–21.
- [3] WANG S, ZHANG J, ZONG C Q. Exploiting word internal structures for generic Chinese sentence representation[C]. Proceedings of the Conference on Empirical Methods in Natural Language Processing, Copenhagen, EMNLP, 2017:298–303.
- [4] MOU L, YAN R, LI G, et al. Backward and forward language modeling for constrained sentence generation[J]. Computer Science, 2016, 4(6):473–482.
- [5] YAEL N, DAVID G, YOAV G, et al. Generating haiku with word associations norms[C]. Proceedings of the Workshop on Computational Approaches Linguistic Creativity, 2009:32–39.
- [6] HUGO G O, TIAGO M, ANA B, et al. Co-PoeTryMe: Interactive poetry generation[J]. Cognitive Systems Research, 2019, 54:199–216.
- [7] ZHOU C L, YOU W, DING X. Genetic algorithm and its implementation of automatic generation of Chinese songci[J]. Journal of Software, 2010, 21(3):427–437.
- [8] LONG J, MING Z. Generating Chinese couplets using a statistical MT approach[C]. In Proceedings of the 22nd International Conference on Computational Linguistics, Manchester, UK: AC, 2008:18–22.
- [9] JING H, MING Z, LONG J. Generating Chinese classical poems with statistical machine translation models[C]. In Proceedings of the 26th AAAI Conference on Artificial Intelligence, Toronto, Canada, 2012:22–26.
- [10] ZHANG X, Lapata M. Chinese poetry generation with recurrent neural networks[C]. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, 2014:25–29.
- [11] WAGN Q, LUO T, WANG D. Chinese song lambics generation with neural attention-based model[C]. In: IJCAI'6, New York, 2016:9–15.
- [12] YI X Y, LI R Y, SUN M S. Generating Chinese classical poems with RNN encoder-decoder[C]. Chinese Computational Linguistics Internaional Symposium on Natural Language Processing Based on Naturally Annotated Big Data, The 16th China National Conference, Nanjing, CCL, 2017:13–15.
- [13] WANG Z, HE W, WU H, et al. Chinese poetry generation with planning based neural network[C]. The 26th International Conference on Computational Linguistics, Osaka, Japan, 2016:1051–1060.
- [14] YU L, ZHANG W, WANG J, et al. Sequence generative adversarial nets with policy gradient[C]. Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, California, USA, 2017:2852–2858.
- [15] ZHE H T, WANG W, CHEN W, et al. Automatic generation of news comments based on gated attention neural networks[J]. IEEE Access, 2018(6):702–710.
- [16] IOANNIS K, SRINIVASAN I, MARK Y, et al. Neural AMR: Sequence-to-sequence models for parsing and generation[C]. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, Vancouver, Canada, 2017:146–157.
- [17] SONG L, ZHANG Y, WANG Z G, et al. A graph-to-sequence model for AMR-to-text generation[C]. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Melbourne, Australia, 2018:1616–1626.
- [18] RIK K, DHANUSH B, LUAN Y, et al. Text generation from knowledge graphs with graph transformers[C]. Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics, 2019:2284–2293.
- [19] DZMITR B, KYUNGHYUN C, YOSHUA B. Neural machine translation by jointly learning to align and translate[C]. CoRR, 2014:1–15.

作者简介:

- 杨泰康(1991–),男,硕士生.研究领域:自然语言处理,信息安全.
- 杨婉霞(1979–),女,博士,副教授.研究领域:信息处理,自动控制.本文通讯作者.
- 刘燕(1987–),女,硕士,讲师.研究领域:信息处理.
- 胡智喻(1995–),男,硕士生.研究领域:自然语言处理.
- 王巧珍(1996–),女,硕士生.研究领域:自然语言处理,信息安全.
- 徐明杰(1998–),男,硕士生.研究领域:自然语言处理.