

# 诗词自动生成隐写算法与系统

芦晶, 赵翔, 张渊皓, 杨婉霞, 周蓓蓓

(甘肃农业大学机电工程学院, 甘肃兰州 730070)

**摘要:** 中国古诗词具有结构工整、情感丰富、篇幅适中等特点, 适合作为生成式无载体信息隐写的天然载体, 但由于其语义含蓄、寓意深远、风格多样, 使得自动生成诗词的隐写极具挑战性。随着大语言生成模型的出现, 诗词生成的质量得以大幅提高, 但将大语言模型应用于诗词生成的隐写还未有报道。为此, 引入BERT情感分析模型, 结合Attention机制设计了Seq2Seq自动生成诗词的隐写算法和模型, 并基于PN40构建了相应的硬件系统和GUI界面。在主题词引导以及格律/情感/互信息的约束下, 生成多模式隐写诗词并实现了系统上的快速输出。实验结果表明, 所提模型生成的隐写诗词主题明确, 情感一致, BLEU评测值高达44.3%, 情感分析平均准确率均高于85%, 极大增强了隐写诗词的感知和统计隐蔽性, 加快了生成式信息隐写的应用。

**关键词:** 诗词; 大语言模型; 情感分析; 隐蔽性; 信息隐写

DOI: 10.11907/rjtk.241279

中图分类号: TP181

文献标识码: A

开放科学(资源服务)标识码(OSID):

文章编号: 1672-7800(2024)009-0048-08



## Steganography Model and System for Automatic Generated Poem

LU Jing, ZHAO Xiang, ZHANG Yuanhao, YANG Wanxia, ZHOU Beibei

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

**Abstract:** Chinese ancient poetry has the characteristics of neat structure, rich emotions, and appropriate length, making it suitable as a natural carrier for generative information hiding. However, due to its implicit semantics, profound meaning, and diverse styles, the automatic generation of poetry hiding is extremely challenging. With the emergence of big language generation models, the quality of poetry generation has been greatly improved, but there have been no reports on applying big language models to implicit writing in poetry generation. To this end, the BERT sentiment analysis model was introduced and combined with the Attention mechanism to design a hidden writing algorithm and model for Seq2Seq to automatically generate poetry. The corresponding hardware system and GUI interface were constructed based on PN40. Under the guidance of theme words and the constraints of rhythm/emotion/mutual information, multi-mode implicit poetry was generated and quickly output on the system. The experimental results show that the proposed model generates implicit poems with clear themes and consistent emotions. The BLEU evaluation value is as high as 44.3%, and the average accuracy of sentiment analysis is above 85%, greatly enhancing the perception and statistical concealment of implicit poems and accelerating the application of generative information steganography.

**Key Words:** poetry; large language model; sentiment analysis; concealability; information steganography

## 0 引言

古诗词结构工整、情感丰富、篇幅适中, 成为无载体信

息隐写的最佳选择之一<sup>[1]</sup>。到目前为止, 诗词自动创作已在语料库建立、语义分析、自动生成等方面取得了一定的研究成果<sup>[2-4]</sup>。然而, 隐写诗词的生成依然存在主题和情感一致性的挑战。随着深度学习技术在自然语言处理方

收稿日期: 2024-03-20

扫描二维码阅读全文:

基金项目: 国家自然科学基金项目(62262002, 61862002)

**作者简介:** 芦晶(1998-), 女, 甘肃农业大学机电工程学院硕士研究生, 研究方向为自然语言处理和知识图谱; 赵翔(2000-), 男, 甘肃农业大学机电工程学院硕士研究生, 研究方向为自然语言处理、文本生成; 张渊皓(2000-), 男, 甘肃农业大学机电工程学院硕士研究生, 研究方向为自然语言处理、情感分析; 杨婉霞(1979-), 女, 博士, 甘肃农业大学机电工程学院教授、博士生导师, 研究方向为自然语言处理、文本生成; 周蓓蓓(1978-), 女, 硕士, 甘肃农业大学机电工程学院讲师, 研究方向为信号处理。本文通讯作者: 杨婉霞。



面的广泛应用,循环神经网络(Recurrent Neural Network, RNN)、Transformer等模型在文本生成领域的应用不断改进和优化,一定程度上提高了隐写诗词的生成质量<sup>[5-6]</sup>。然而RNN存在梯度消失问题,对于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问题,但不论输入长短,它都会将输入编码成一个固定长度的向量表示,使得模型对长输入序列的学习效果依然较差。Transformer存在长程依赖建立能力差、时间维度建模能力欠缺的问题。

在诗词隐写领域,文献[7]最先提出并实现了基于宋词生成的信息隐藏方法,但实现该算法对应的词典容量较小,生成的隐写宋词可读性弱,容易引起攻击者的注意,且嵌入率较低;文献[8]提出一种基于宋诗的文本隐写算法,嵌入率达到27.1%,但生成隐写宋词的主题不够明确,语义关系还有待提高;文献[9]基于马尔可夫模型设计了宋词生成隐藏算法,但生成的宋词主题发散,嵌入率较低;文献[10-11]通过设置候选词的方式将信息嵌入长短期记忆网络(Long Short-Term Memory, LSTM),在提高嵌入率的同时保证了生成文本的可读性;文献[12]通过构建意象词汇秘密词典设计了基于唐诗智能生成的信息隐藏算法,在隐藏容量和隐蔽性方面取得良好效果;文献[13-14]提出基于绝句和宋词的生成构造式信息隐藏算法,尽可能地应用诗词的结构特征实现秘密信息的隐藏,增强了隐写诗词的可读性和隐藏容量,但生成的诗词风格较为单一,质量有待提高;文献[15-16]通过检索宋词韵律“中平仄”出现频率的奇偶性实现信息隐藏,但其隐藏容量较低;文献[17]通过格律、互信息等约束构造候选词库实现宋词生成的信息隐藏,提高了隐写容量,但生成的隐写宋词情感无约束。

由上述研究可知,现有自动生成隐写诗词的方法基本都考虑了主题和结构要求,但对生成诗词的情感毫无约束,使得生成的隐写诗句前后情感可能不一致,将降低隐写诗词的感知隐蔽性,影响其安全性。尽管大语言模型生成的文本及诗词质量较好,但是目前尚未有该类模型生成隐写文本的报道。为此,本文引入情感分析模型,设计了一种基于BERT(Bidirectional Encoder Representations from Transformers)预训练模型结合Attention机制的Seq2Seq自动生成诗词的信息隐写模型和算法,对生成隐写诗句的主题和情感均加以约束,使得整首诗词的主题和情感前后一致,增强了诗词隐写生成的隐蔽性。同时基于PN40实现了相应的隐写系统,以便于实际操作与应用。

## 1 诗词生成隐写模型

本文设计的诗词生成隐写模型结构如图1所示,主要思想是在主题/格律/情感约束下生成诗词并嵌入私密信息。其运行流程为:首先,用户将创作意图以句子或词语的形式输入,并选择相应的功能,系统对输入信息进行主

题词的提取和扩展;其次,通过BERT词嵌入模型将相应的主题词及已有句子转换为词向量表示;最后,利用主题词、已生成诗句、格律模板和情感约束的方式生成完整的诗词。在进行诗词隐写时,将用户输入的私密信息转换成二进制比特流,在主题词选取和基于互信息约束的诗词生成过程中进行嵌入,实现诗词自动生成过程的信息隐藏。

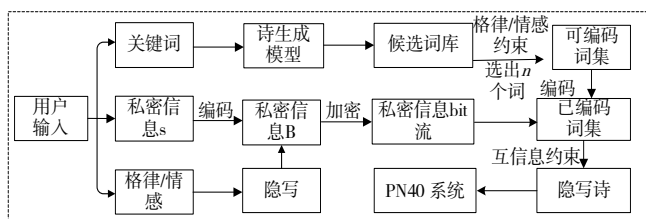


Fig. 1 Steganographic model of poetry generation

图1 诗词生成隐写模型

## 2 诗词生成模型构建

### 2.1 主题规划与情感分析

在诗词创作过程中,主题与情感的表达至关重要,也是诗词自动生成的重要挑战。为解决单个主题词引导整首诗词生成的主题偏移问题,本文引入主题规划方式,采用提取(TextRank算法)或扩展(BERT词向量空间)的主题词对每个句子生成的主题加以控制,从而使整首诗主题明确。同时采用BERT模型分析作者所要表达的情感,通过构建情感词典对生成诗句的情感加以约束,以保证其一致性。

采用情感语料对BERT情感分析模型进行充分训练,以实现对本情感的正确分类<sup>[18]</sup>。具体步骤为:首先,将收集的诗词数据整理为“语句+标签”的格式(即句子片段索引序列 token\_ids, 句子标签 label, 句子长度 seq\_len, 句子中的无效位置 mask)作为BERT模型的输入;其次,将输入数据划分为batch进行迭代,利用交叉熵损失函数(Cross-Entropy Loss Function)衡量误差,在自适应动量估计(Adaptive Moment Estimation, Adam)优化器的调节下得到最优分类模型;最后,将输入样本对应各标签的概率值从BERT最后一层输出,从中选取最大值作为模型的预测结果。

### 2.2 格律与情感双约束

与其他文学体裁相比,诗词具有鲜明的结构特征,即依据韵律特性形成的格律。例如,“风急天高猿啸哀,渚清沙白鸟飞回”的格律为仄仄/平平/仄仄平,平平/仄仄/仄平平。为此,本文提出格律与情感双约束的方法指导诗词生成,即按照选定的格律模板将诗词语句分割成对应的韵律块。在生成每个韵律块中的字符时,首先按照其概率大小进行排序;其次基于平仄词典和情感词典(积极/消极词典)剔除不符合诗句平仄要求和情感要求的字词,形成新的候选词库,从中选词以生成完整诗句。本文构建的平仄词典和情感词典如图2所示,其中平声为1、仄声为2;积极词为1、消极词为2。



Fig. 2 Examples of tonal and emotional dictionaries

图2 平仄词典和情感词典示例

### 2.3 诗词生成模型

基于上述思想构建基于Seq2Seq的诗词生成模型,具体结构如图3所示。其中,词嵌入部分采用BERT预训练语言模型,以获得词和句子动态而丰富的上下文语义信息。通过输入诗词语句和字典对BERT词嵌入模型进行预训练后可以输出全面准确表示诗词语义的向量信息,以便作为编码器的输入。编码器采用BiLSTM(Bi-directional Long Short-Term Memory)网络对诗词的输入序列进行双向编码和特征提取,以获得更多上下文语义信息。Attention机制通过判断编解码两侧隐藏状态 $H_i$ 和 $S_i$ 的相关性计算出注意力权重 $A_w$ ,将 $A_w$ 与Encoder端的隐藏状态加权求和,得到输入的上下文信息 $C_i$ ;然后将 $C_i$ 与解码器端的隐藏状态 $S_i$ 融合用于解码,输出诗词句子的字向 $(y_0, y_1, \dots, y_t)$ ,通过向量转换将其映射到词表中。在此,设计集束搜索(Beam Search)算法筛选出最符合语义的字符。模型部分计算公式为:

$$A_w = \frac{\exp(\text{score}(H_i, S_i))}{\sum_{i=1}^t \exp(\text{score}(H_i, S_i))} \quad (1)$$

$$C_i = \sum_{i=1}^t A_w \cdot H_i \quad (2)$$

$$S_i = g(y_{i-1}, S_{i-1}, c_i) \quad (3)$$

$$y_i = W_0 S_i + b_0 \quad (4)$$

式中: $H_i$ 为编码器端的隐藏层状态, $S_i$ 为解码器端第 $i$ 个字符的隐藏状态, $y_i$ 为输出的第 $i$ 个字符, $g()$ 为非线性激活函数。

## 3 信息隐藏与提取方法

### 3.1 信息隐藏算法

首先使用Unicode对输入的秘密信息进行编码与加密,形成秘密信息比特流 $B$ ;然后利用训练好的诗词生成模型以及用户输入的主题词、格律要求生成一首诗/词 $Y$ ,接着利用诗/词句子中每个字符位置候选字的概率分布建立候选词集。在格律和情感的双重约束下,选择前 $n$ 个字

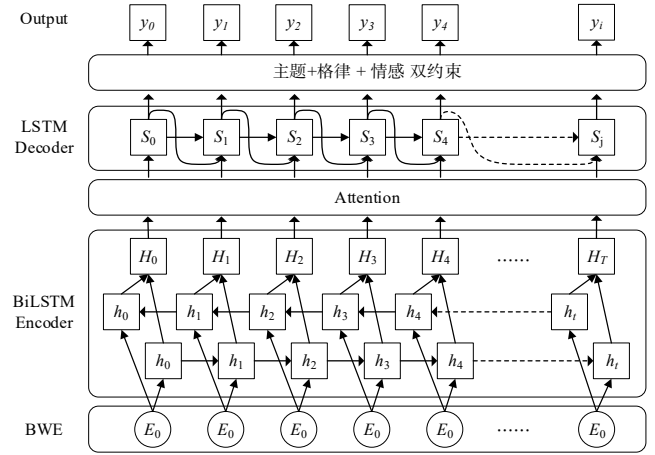


Fig. 3 Poetry generation model structure

图3 诗词生成模型结构

形成可编码词集,并以此建立霍夫曼树进行霍夫曼编码,得到已编码词集;最后从秘密信息比特流 $B$ 中选择字符,将信息嵌入。如果单首诗词不能完成全部秘密信息嵌入,则会进行提示,并再次按照相同的格律和主题要求生成下一首诗词,直至秘密信息全部嵌入为止。本文提出的信息隐藏算法伪代码为:

输入:词牌(Tone);主题(Theme):得到 $[Word = Word_1, Word_2, \dots, Word_k]$ 和 $Sentiment=P/M/N$ ;词库大小(Size):可编码主题词集 $Word\_Size=m$ ,可编码词集 $Return\_Size=n$ ;将秘密信息( $S$ ="今晚集合")进行编码加密得到二进制比特流: $B=[0, 1, 0, 0, 1, 1, \dots]$

输出:完整的隐写诗词 $P$

1. 数据集的预处理
2. 诗词生成模型的训练
3. 根据训练好的诗词生成模型和各参数生成诗词 $Y = (y_1, y_2, \dots, y_n)^T$
4. 根据参数需求,将秘密信息划分成 $B_1$ 和 $B_2$
5. FOR(不是秘密信息 $B_1$ 的结尾)
6. 提取并扩展关键词,形成候选主题词库
7. FOR(遍历Word):
8. 对候选主题词库按相似度降序排列,选择前 $m$ 个主题词构成可编码主题词集,并进行二进制升序编码
9. 按照秘密信息选择对应的主题词,并删除该词后,重新降序排列
10. RETURN 主题词列表
11. FOR(不是秘密信息 $B_2$ 的结尾)
12. FOR(遍历诗词 $Y$ 的每一句)
13. WHILE(遍历当前诗句 $y_i$ 的每个字,并删除之前用过的字)
14. IF(是当前诗句的句首)
15. 将该位置字的概率降序排列,经平仄词典和情感词典筛选出不符合韵律和情感要求的字,选择前 $n$ 个字作为可编码词集
16. 对可编码词集建立霍夫曼树,按照左侧为0,右侧为1对其进行编码

17. 读取加密后的秘密信息,选择霍夫曼编码词集中相应的字  $w_p$
18. IF(是当前诗句同一韵律块中的首字)
19. 同第14、15、16步
20. ELSE
21. 计算  $w_p$  与上一位置候选字互信息,然后按互信息对候选字降序排列
22. 同第14、15、16步
23. RETURN 生成的诗词

### 3.2 秘密信息提取

秘密信息提取与隐藏过程相反。首先对含密诗词进行分析,了解其格式、韵律;然后利用发送方共享的主题语句、主题词列表、相应的词库和模型等进行诗词生成,其中主题词扩展和诗词生成过程同样会产生候选词集,将该词集进行编码后与隐写诗词进行对照,分别从中提取出隐藏信息的比特流  $B_1$  和  $B_2$ ;最后将  $B_1$  与  $B_2$  拼接起来便可得到完整的秘密信息比特流  $B$ ,对其进行解密解码可得到隐藏的秘密信息。秘密信息提取算法的伪代码如下:

输入:主题语句(*Theme*)和主题词列表(*Word\_list*),得出  $Sentiment=P/M/N$ ;完整的含密诗词  $P$

输出:秘密信息( $S$ ="今晚集合")

1. 对主题语句进行主题词提取和扩展,并选择前  $m$  个主题词构成可编码主题词集,进行二进制升序编码
2. FOR(遍历 *Word*)
3. 对候选主题词库按相似度降序排列,选择前  $m$  个主题词构成可编码主题词集,并进行二进制升序编码
4. 对照已知的主题词提取出比特流,并添加到  $B_1$  中
5. 删除该词后,重新降序排列
6. RETURN 提取的二进制比特流  $B_1$
7. 根据训练好的诗词生成模型和相关参数生成诗词  $Y = (y_1, y_2, \dots, y_N)^T$
8. FOR(遍历生成隐写诗词的每个句子)
9. FOR(遍历诗词  $Y$  的每一句)
10. WHILE(遍历当前诗词语句的每个字,并删除之前用过的字)
11. IF(是当前诗词语句的句首)
12. 将该位置字的概率降序排列,经平仄词典和情感词典筛选出不符合韵律和情感要求的字,选择前  $n$  个字作为可编码词集
13. 对可编码词集建立霍夫曼树,按照“左0右1”对其进行编码
14. IF(下一时刻字在候选词集中)
15. 由当前时刻的字,确定从叶节点到根节点的路径
16. 按照霍夫曼树的编码规则,读出二进制的比特流
17. 输出解码的比特流并添加到  $B_2$  中
18. ELSE
19. 初次解码过程结束
20. ELSE

21. 计算韵律块首字与该位置候选字的互信息,然后按互信息降序排列
22. 同第11、12步
23. 同第13步
24. 同第14、15、16步
25. 同第17步
26. 同第18步
27. RETURN 提取的二进制比特流  $B_2$
28. 对提取的秘密信息比特流  $B = B_1 + B_2$  进行解密,解码

## 4 实验方法与结果分析

### 4.1 实验数据集

本文使用的诗词语料主要从互联网以及中文诗歌古典文集数据库(Chinese-Poetry)中收集而得,重点针对绝句、律诗和各朝代的词进行整理,得到17万余首古诗和5万余首词,几乎涵盖了所有诗词语料。

将题目、作者等信息从诗词中分离,得到只包含诗句的数据集。根据停用词库去除停用词等噪音数据,生成含有8000个字的诗词字典。对数据集中的句子进行分词并提取主题词,得到包含诗句和主题词的两个数据集。按照同样的方式对词的词牌等信息进行预处理。对数据集进行预处理后按照8:1:1的比例分为训练集、验证集和测试集。具体信息如表1所示。

Table 1 Dataset information

表1 数据集信息

数据	分类	数目/首	句子数/条
诗	五言绝句	训练集	13 011
		验证集	1 500
	五言律诗	测试集	1 500
		训练集	40 233
		验证集	5 000
		测试集	5 000
七言绝句	训练集	51 385	
	验证集	6 000	
	测试集	6 000	
	训练集	42 366	
七言律诗	验证集	5 000	
	测试集	5 000	
	训练集	43 483	
	验证集	5 000	
词	验证集	5 000	
	测试集	5 000	

### 4.2 模型训练方法

#### 4.2.1 情感分析模型训练

该模型的训练数据为通过互联网收集到的大量已标注数据集,整理后得到白话文和古诗词两个部分,按照8:1:1的比例分为训练集、验证集和测试集。具体如表2所示。

#### 4.2.2 诗词生成模型训练

诗词生成模型的实质为输入空间  $[X_1, X_2, \dots, X_n]$  到输出空间  $[Y_1, Y_2, \dots, Y_n]$  的真实映射函数  $Y_n = f(X_n)$ , 简单来

Table 2 Emotion analysis model training dataset

表2 情感分析模型训练数据集

数据	分类	数目/条
中文白话文(55 324句); 古诗词语句(20 000句)	积极(P)	22 335
	中性(M)	30 532
	消极(N)	22 457
	训练集	60 324 (P:18 462,M:23 351,N:18 511)
	验证集	7 500 (P:1 922,M:3 525,N:1 983)
测试集	7 500 (P:1 951,M:3 656,N:1 963)	

说就是表征输入词向量到输出诗词语句向量的真实条件概率分布。其训练过程即为函数中各权重的更新过程,目的是使函数经验风险最小化(Empirical Risk Minimization, ERM)。计算公式为:

$$l(\theta) = -\frac{1}{N} \sum_{i=1}^N \log P(Y_i|X_i, \theta) \quad (5)$$

式中: $X_i$ 和 $Y_i$ 分别为输入序列 $[X_1, X_2, \dots, X_n]$ 和输出的诗句序列 $[Y_1, Y_2, \dots, Y_n]$ ;  $\theta$ 表示训练时需要更新的权重参数。简而言之,模型的训练目标就是寻找一组参数 $\theta$ 得到ERM,使模型预测值与真实值间相差最小。

### 4.3 参数设置

超参数(Hyperparameter)是模型能否达到最优效果的关键因素。本文模型超参数设置如表3所示。选择高效的Adam算法作为优化器,采用真值(Ground\_truth)方式和计划采样(Scheduled\_sampling)两种方式训练模型。

Table 3 Model hyperparameter settings

表3 模型超参数设置

参数	值
Embedding size	128
Hidden size	128
Each encoder and decoder layers	4
Learning_rate	0.000 2
Dropout_rate	0.3
Batch size	64
Max epochs	1 000

为了体现Attention机制的优势,并探究不同神经网络编码单元、不同训练方式对生成模型性能的影响,本文从LSTM和GRU两种神经网络、Bahdanau和Luong两种Attention机制、Ground\_truth和Scheduled\_sampling(sampling\_probability=0.1)两种训练方式3个方面设置LSTM+Bahdanau+ground\_truth(Model\_L\_BG)、GRU+Bahdanau+Ground\_truth(Model\_G\_BG)、LSTM+Luong+ground\_truth(Model\_L\_LG)、LSTM+Bahdanau+Scheduled\_sampling(Model\_L\_BS)、LSTM+ground\_truth(Model\_L\_\*G)5组模型分别进行实验,选取最优模型进行后续实验。各模型具体参数设置如表4所示。

### 4.4 模型评估

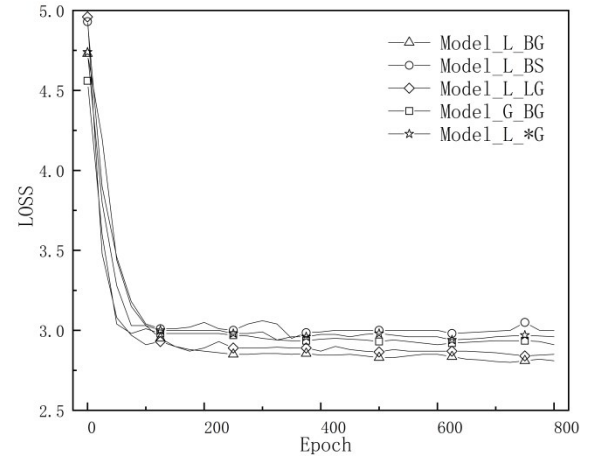
在训练诗词生成模型时,为了得到最优模型,分析比较同为BERT词嵌入的5组参数训练模型的Loss值和Perplexity值随Epoch的变化情况。实验结果如图4所示。可以看出,GRU网络参数量较LSTM更少,能够较快实现收

Table 4 Model parameter settings of comparison experiment

表4 对照实验模型参数设置

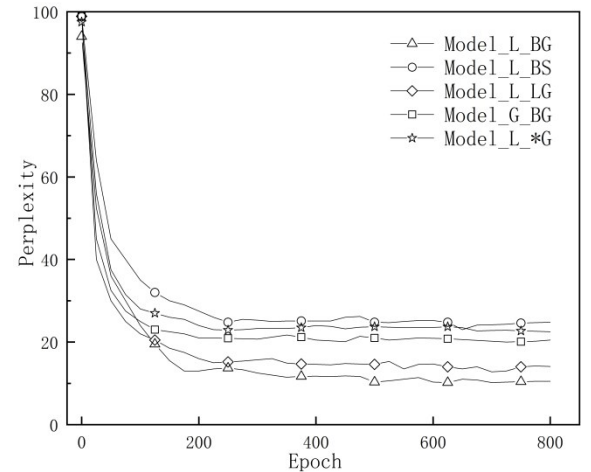
模型	神经网络	Attention 机制	训练方式
Model_L_BG	LSTM	Bahdanau	Ground_truth
Model_L_BS	LSTM	Bahdanau	Scheduled_sampling
Model_L_LG	LSTM	Luong	Ground_truth
Model_G_BG	GRU	Bahdanau	Ground_truth
Model_L_*G	LSTM	/	Ground_truth

敛,但当数据量较大时,LSTM网络能够更加全面地提取特征,编码效果更佳。对于短文本的生成,Bahdanau Attention相较Luong Attention更加稳定,且较无Attention机制表现更优。在诗词类短文本生成中,由于序列并不长,Ground\_truth的训练方式较Scheduled\_sampling表现更优。实验设置的5个模型虽然均能较好地收敛,但从Loss值和Perplexity值综合来看,Model\_L\_BG表现最优。



(a) The change of Loss value with the Epoch

(a) Loss随Epoch的变化



(b) The change of perplexity with the Epoch

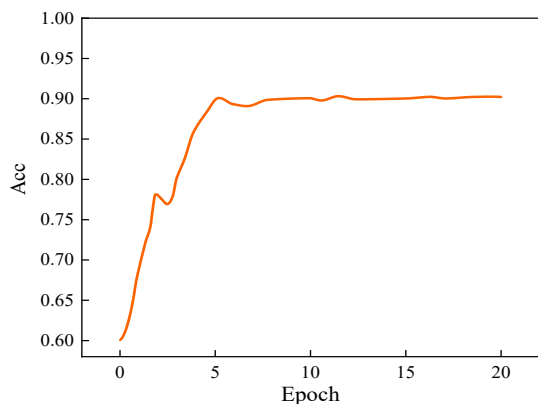
(b) Perplexity随Epoch的变化

Fig. 4 Training effect of models with different parameters

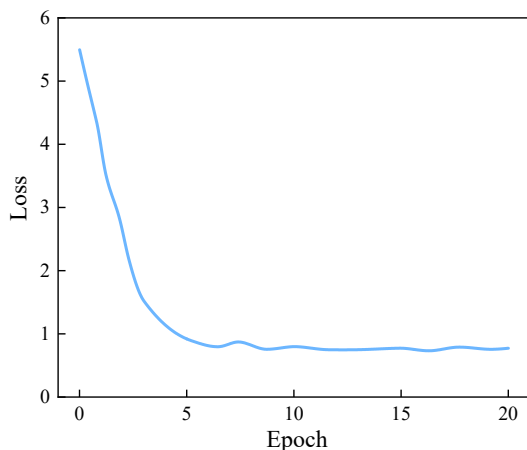
图4 不同参数模型训练效果

本文构建的基于BERT的诗词情感分析模型损失率和准确率随Epoch的变化情况如图5所示。可以看出,该模型微调方便快捷,Loss值能较好地收敛,准确率高达0.91,

可实现对文本语句情感的准确分类。



(a) Accuracy  
(a) 准确率



(b) Loss rate  
(b) 损失率

Fig. 5 The loss and accuracy of sentiment analysis model

图 5 情感分析模型损失率与准确率

#### 4.5 生成诗词质量测评

采用 BLEU (Bilingual Evaluation Understudy)<sup>[20]</sup> 对 BERT+Model\_L\_BG、Word2vec+Model\_L\_BG、BERT+Model\_L\_\*G 和 Model\_L\_BG 4 种模型生成诗词的质量进行评估。具体操作为:首先选择相似主题原语料的五言绝句、七言绝句、五言律诗、七言律诗、宋词各 20 首,人工从这 5 类原诗词中选取主题词输入各自训练好的模型中;然后使用束搜索方式进行预测,在格律和情感的双重约束下生成各类诗词;最后计算并比较每组主题生成诗词的 BLEU 平均值。测评结果如表 5 和图 6 所示。

可以看出,使用 BERT 词嵌入生成的诗词在 BLEU 测

Table 5 BLEU test results

表 5 BLEU 测评结果

模型	五言绝句	五言律诗	七言绝句	七言律诗	宋词
BERT+Model_L_BG	0.425	0.439	0.434	0.442	0.406
Word2vec+Model_L_BG	0.347	0.356	0.349	0.358	0.323
Bert2+Model_L_*G	0.361	0.374	0.373	0.376	0.352
Model_L_BG	0.328	0.336	0.334	0.339	0.317

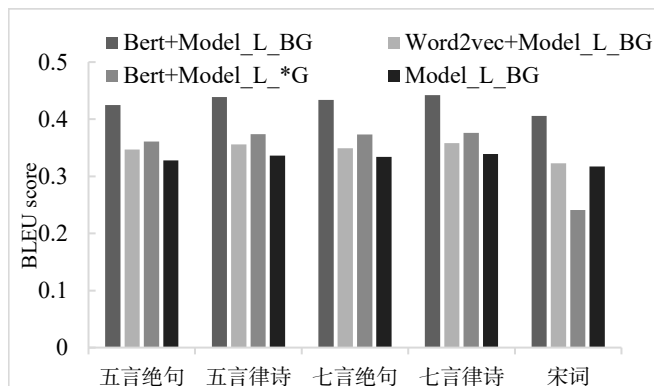


Fig. 6 BLEU test results

图 6 BLEU 测评结果

评中优于 Word2vec 词嵌入模型生成的诗词,且生成的七言绝句质量较其他类别更优。从 BERT+Model\_L\_BG 和 BERT+Model\_L\_\*G 模型生成诗词的测评结果可知,使用 Attention 机制后能明显改善生成诗词的质量。而模型 Model\_L\_BG 在加入 Word2vec 后,5 种类型诗词生成的 BLEU 均有所提升。综合来看,律诗因语料数量更多、语义更加丰富,整体生成效果优于绝句;而宋词由于格律的复杂性,结果逊色于古诗。

为进一步探究所提情感约束方法的有效性,对生成诗词的情感进行评估。具体操作为:首先选择 100 条情感不一的主题词句,采用基于 BERT 的情感分析模型对其进行分类;然后将主题词句分别输入基于 BERT+Model\_L\_BG 训练的五言绝句、七言绝句、五言律诗、七言律诗、宋词模型中,在情感词典的约束下生成相应的诗词;接着利用基于 BERT 的情感分析模型对生成的每句诗词进行情感分类;最后将诗词各句的主题情感和输入的主题情感进行对比,计算其平均准确率。测评结果如图 7 所示。可以看出,基于 BERT 情感分析模型和情感词典的情感约束方式能较好地约束生成诗词的情感,各类诗词情感分析的平均准确率均高于 85%,可有效表达创作者的情感诉求。

## 5 基于 PN40 的系统设计

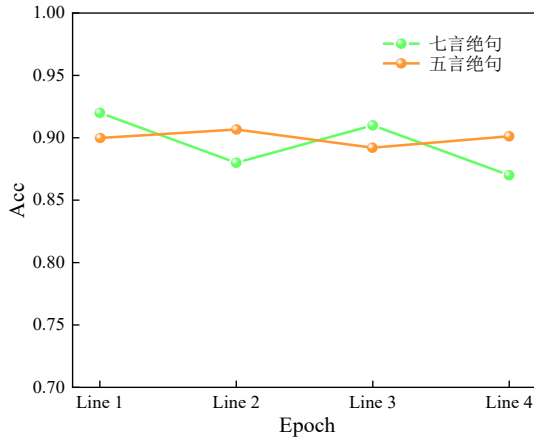
### 5.1 系统设计

PN40 硬件系统主要由输入模块、主机模块和输出显示模块 3 个部分组成。为实现诗词自动生成与隐写的诸多功能,首先在 Vmware 平台搭建虚拟机,基于 PyQt5 进行 GUI 界面的设计,主要实现系统的交互功能,并显示返回的结果,其架构如图 8 所示;然后基于 PN40 设计硬件载体,构建 Ubuntu 系统及其深度学习环境;最后将相关程序迁移至硬件载体。

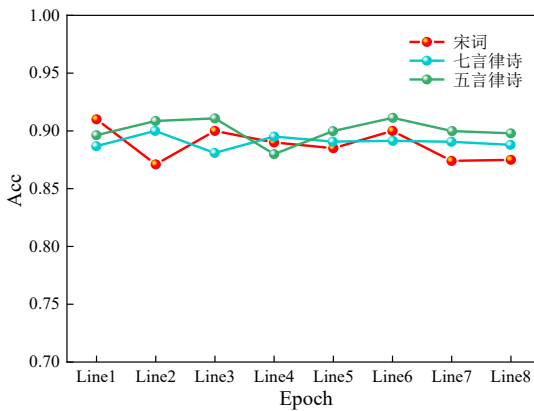
主要硬件参数和相关环境设置分别如表 6 和表 7 所示。

### 5.2 系统时效性评估

目前已有的诗词创作大多存在低时效问题,往往需要



(a) Average accuracy 1  
(a) 平均准确率 1



(b) Average accuracy 2  
(b) 平均准确率 2

Fig. 7 Generated poetry assessment results

图7 生成诗词测评结果

排队等候较长时间,而且质量参差不齐,功能相对单一。本文构建的诗词自动生成与隐写系统集五言/七言、平韵/仄韵、绝句/律诗、藏头诗、各词牌的词等多类型于一体,且从输入信息到输出结果耗时较短,极大地缩短了创作时间。时效性测评结果如表8所示。此外,该系统首次将诗词生成与隐写相结合,方便用户更好地理解生成式无载体

Table 6 Main hardware parameters

表6 主要硬件参数

参数	设置
CPU	Intel(R) Celeron(R) J4125@2.00GHz
GPU	UHD Graphics 600
RAM	4GB SODIMM DDR4 Synchronous 2400MHz (0.4 ns) × 2
ROM	Lexar 128GB SSD
Interface	USB3.1/2.0、VGA Type-C、HDMI
Features	WIFI、Bluetooth、Wired network、audio
LED	ZGYNK(R) HD LED MONITOR
PN40 Size	115×115×49mm
Screen Size	155×87mm

Table 7 Environment settings

表7 环境设置

参数	设置
System	Ubuntu 18.04.2 LTS
Python	2.7
Tensorflow	1.14
Keras	2.25
Jieba	0.38
Gensim	2.0
PyQt	5.8

信息隐藏技术。

Table 8 System timeliness evaluation result

表8 系统时效性测评结果

时效度量	具体内容	预计耗时/s
启动系统	输入命令启动系统,直到GUI界面显示	5~7
信息输入	输入主题词句等信息	5~9
功能选择	选择各项功能指标	5~8
诗词生成与隐写	点击生成,直到输出结果	5~7

## 6 结语

结合诗词语言精炼、语义情感丰富的特点,本文以诗词为载体设计了主题和情感可控的诗词自动生成隐写模型,并将此模型迁移至基于PN40的系统,实现了隐写模型快速、便捷和可视化应用。实验结果表明,所提诗词自动生成隐写算法和模型生成的诗词主题一致,情感可控,

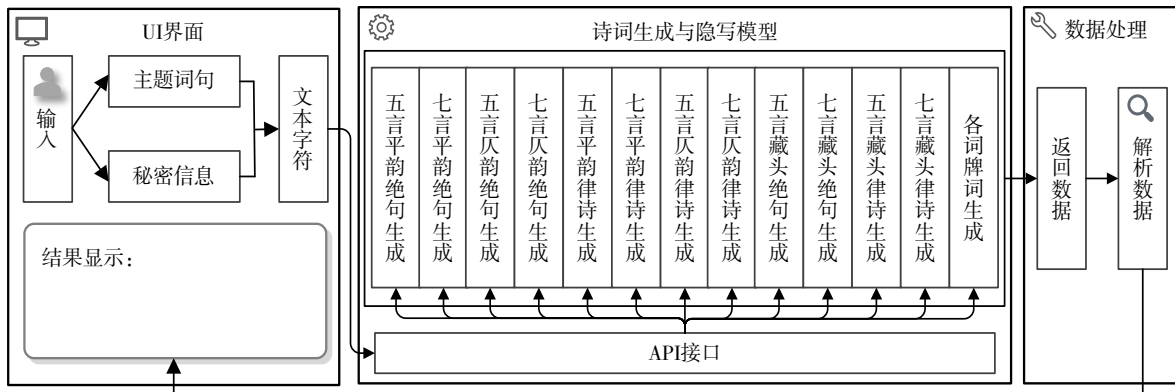


Fig. 8 GUI interface design of poetry generation steganography system

图8 诗词生成隐写系统的GUI界面设计

BLEU测评值达到了44.2%;诗词生成及隐写系统的实效性高,平均运行时间不到10s,操作便捷,可实现性好。后续可从情感细腻化和语义丰富化方面进一步提升模型性能,从编码和隐写效率方面改进隐写算法。

#### 参考文献:

- [1] HU Z Y, YANG W X, YANG K T, et al. Automatic generation of Song Ci combining BERT word embedding and attention mechanism[J]. Software Guide, 2021, 20(11): 1-9.  
胡智喻,杨婉霞,杨泰康,等. 结合BERT词嵌入与注意力机制的宋词自动生成[J]. 软件导刊, 2021, 20(11): 1-9.
- [2] LIANG J N, SUN M S, YI X Y, et al. Poetry retrieval based on deep neural network[J]. Journal of Chinese Information Processing, 2020, 34(12): 82-91.  
梁健楠,孙茂松,吴晓沅,等. 基于深度神经网络的诗词检索[J]. 中文信息学报, 2020, 34(12): 82-91.
- [3] ZHANG S Y, YANG Z L, YANG J S, et al. Linguistic steganography: from symbolic space to semantic space[J]. IEEE Signal Processing Letters, 2021, 28: 11-15.
- [4] YANG Z L, ZHANG S Y, HU Y T, et al. VAE-Stega: linguistic steganography based on variational auto-encoder[J]. IEEE Transactions on Information Forensics and Security, 2020, 16: 880-895.
- [5] XIANG L Y, YANG S H, WANG R, et al. Constrained natural language information hiding method based on sequence to steganographic sequence[J]. Chinese Journal of Computers, 2023, 46(8): 1650-1669.  
向凌云,杨双辉,王蓉,等. 基于序列到隐写序列的约束型自然语言信息隐藏方法[J]. 计算机学报, 2023, 46(8): 1650-1669.
- [6] PENG W L, WANG T, QIAN Z X, et al. Cross-modal text steganography against synonym substitution-based text attack[J]. IEEE Signal Processing Letters, 2023, 30: 299-303.
- [7] YU Z S, HUANG L S, CHEN Z L, et al. Using Song Ci to realize text information hiding with high embedding rate[J]. Journal of Chinese Information Processing, 2009, 23(4): 55-63.  
余振山,黄刘生,陈志立,等. 用宋词实现高嵌入率文本信息隐藏[J]. 中文信息学报, 2009, 23(4): 55-63.
- [8] LIU Y C, WANG J, WANG Z B, et al. A technique of high embedding rate text steganography based on whole poetry of Song dynasty[C]//Nanjing: International Conference on Cloud Computing and Security, 2016.
- [9] LUO Y B, HUANG Y F, LI F F, et al. Text steganography based on ci-poetry generation using Markov chain model[J]. KSII Transactions on Internet and Information Systems, 2016, 10(9): 4568-4584.
- [10] CAO Y, ZHOU Z L, CHAKRABORTY C, et al. Generative steganography based on long readable text generation[J]. IEEE Transactions on Computational Social Systems, 2022, 99: 1-11.
- [11] XIANG L Y, WANG R, YANG Z L, et al. Generative linguistic steganography: a comprehensive review[J]. KSII Transactions on Internet and Information Systems, 2022, 16(3): 986-1005.
- [12] LI B B. Intelligently generated information hiding based on Tang poetry[D]. Guangzhou: Guangzhou University, 2018.  
李彬彬. 基于唐诗智能生成的信息隐藏[D]. 广州: 广州大学, 2018.
- [13] QIN C, WANG M, SI G W, et al. Constructional information hiding algorithm based on the generation of quatrain[J]. Chinese Journal of Computers, 2021, 44(4): 773-785.  
秦川,王萌,司广文,等. 基于绝句生成的构造式信息隐藏算法[J]. 计算机学报, 2021, 44(4): 773-785.
- [14] QIN C, LI R S, QIAN Z X, et al. Large capacity structural information hiding algorithm based on Song Ci generation[J]. Chinese Journal of Computers, 2023, 46(1): 17-30.  
秦川,李蓉受,钱振兴,等. 基于宋词生成的大容量构造式信息隐藏算法[J]. 计算机学报, 2023, 46(1): 17-30.
- [15] ZOU Z Y, WANG K X. Steganography of carrier-free text based on rhyme of Song Ci[J]. Journal of Qingdao University (Natural Science Edition), 2021, 34(1): 7-12.  
邹孜逸,王开西. 基于宋词韵律的无载体文本隐写方法[J]. 青岛大学学报(自然科学版), 2021, 34(1): 7-12.
- [16] WU G H, GONG L C, YUAN L F, et al. Review of information hiding on Chinese text[J]. Journal on Communications, 2019, 40(9): 145-156.  
吴国华,龚礼春,袁理锋,等. 中文文本信息隐藏研究进展[J]. 通信学报, 2019, 40(9): 145-156.
- [17] YANG W X, CHEN S, GUAN L, et al. Song Ci automatically generated information hiding algorithm[J]. Netinfo Security, 2022, 22(7): 46-54.  
杨婉霞,陈帅,管磊,等. 宋词自动生成的信息隐藏算法[J]. 信息安全, 2022, 22(7): 46-54.
- [18] JIA K L. Sentiment classification of microblog: a framework based on BERT and CNN with attention mechanism[J]. Computers and Electrical Engineering, 2022, 101: 108032.
- [19] SUTSKEVER I, VINYALS O, LE Q V. Sequence to sequence learning with neural networks[C]//Montreal: Proceedings of the 27th International Conference on Neural Information Processing Systems, 2014.
- [20] SHAO Y Z, SHAO T, WANG M H, et al. A sentiment and style controllable approach for Chinese poetry generation[C]//Queensland: Proceedings of the 30th ACM International Conference on Information and Knowledge Management, 2021.

(责任编辑:尹晨茹)