

文章编号: 1003-0077(2010)02-0096-08

基于统计的汉语格律诗生成研究

何晶^{1,2}, 周明², 蒋龙²

(1. 清华大学 理论计算机研究中心, 北京 100084;
2. 微软亚洲研究院 自然语言计算组, 北京 100190)

摘要: 古代中文诗歌的巅峰——中文格律诗, 包括律诗和绝句, 是中国古典诗词的奇葩。该文从已有的古今名诗中自动学习作诗知识, 实现了一个中文格律诗的自动生成系统。该系统接收用户选择的表达其思路的若干个关键词作为输入, 首先, 利用相关词汇数据库和语言模型, 实现了根据用户选定的关键词自动生成诗歌的第一句。其次, 我们独创性地将格律诗的上下句关系映射为源语言到目标语言的翻译关系, 设计了一个基于短语的统计机器翻译模型, 从而把诗歌的第 N-1 句作为输入用以生成第 N 句。并提供了一个用户交互式的系统, 使得用户可以在每一步都选择一个最佳诗句。最后, 我们还精心设计了一套翔实的格律诗评测标准, 并通过单句实验和全诗实验证明, 该方法是诗歌产生的一个较好的方法。

关键词: 人工智能; 机器翻译; 统计机器翻译; 诗歌生成; 绝句评测

中图分类号: TP391 **文献标识码:** A

Generating Chinese Metrical Poetry by a Statistical MT Approach

HE Jing^{1,2}, ZHOU Ming², JIANG Long²

(1. Institute for Theoretical Computer Science, Tsinghua University, Beijing 100084, China;
2. Microsoft Research Asia Sigma Center, Beijing 100190, China)

Abstract: Automatic poetry generation is considered difficult. In this paper, we propose a novel statistical approach for automatic generation of traditional Chinese metrical poetry from a few user-supplied keywords. A template-based model is used to automatically generate the first sentence of the poem. A phrase-based statistical machine translation model then generates additional sentences one-by-one. With our interactive model, the user can select the best sentence from the system's N-best output at each step. The approach has been evaluated on the generation of quatrains of 5- and 7-character lines. The evaluation metrics for single lines as well as for the whole generated poem suggest that this method is very promising.

Key words: artificial intelligence; machine translation; statistical machine translation; poem generation; poem evaluation

1 引言

中文格律诗作为中国悠久灿烂的古典文学的重要组成部分, 兼有中文的美感和艺术的灵感。然而格律诗的韵律要求十分严格, 给作诗造成了障碍。

本研究的目的是通过研究格律诗生成的内在规律, 利用计算机辅助进行格律诗的自动生成。

对联可以看作是一种特殊的格律诗。微软亚洲研究院自然语言计算组研发了自动对联系统^[1]。

<http://duilian.msra.cn>

收稿日期: 2009-08-30 定稿日期: 2009-10-28

基金项目: 国家自然科学基金资助项目(60553001); 国家自然科学基金重大资助项目(2007CB807900, 2007CB807901)

作者简介: 何晶(1986—), 女, 博士生, 主要研究方向为理论计算机科学, 自然语言处理; 周明(1964—), 男, 博士, 研究员, 博导。主要研究方向为自然语言处理, 多语言搜索; 蒋龙(1981—), 男, 硕士, 副研究员, 主要研究方向为自然语言处理, 网络挖掘。

这套系统采用了统计机器翻译的方法。它将上下联的对应关系建模成机器翻译中源语言句子和目标语言句子的对应关系。我们在这个工作基础上,将格律诗中的上下句关系建模为机器翻译中的源语言句子和目标语言句子的关系。然而,诗中的句对跟对联在格律、意境方面有许多不同,不能单纯地采用对联的技术。

统计机器翻译把翻译的过程看作是一个搜索过程,也就是对一个给定的原文句子,生成多种可能的译文,然后搜索一个在统计意义下最优的翻译结果。本研究中,我们采用了基于短语的统计机器翻译^[2,12]方法。该方法以短语作为翻译的基本单位。以给定绝句的上句来生成下句为例,系统会首先将上句按照各种可能划分为多个短语,然后将每一个短语利用翻译模型翻译为下一句中的短语,最后结合语言模型组合得到最优的若干候选下句。我们还对传统机器翻译解码器进行了修改,使之生成符合韵律要求的下句。

基于统计机器翻译的格律诗生成方法有一个遗憾:无法生成诗的第一句。为了弥补此不足,我们从古籍《诗学含英》中获取了一个诗歌词汇库,并设计了一个结合节奏模板和语言模型的首句生成模型。

本文接下来的内容是这样安排的:第二节,介绍了诗歌生成领域的相关工作;第三节提出了诗歌生成的总体框架;第四和第五节,分别解释了首句生成模型和基于统计机器翻译原理的自动生成诗歌的模式;第六节报告了实验的设计说明和评测结果,并在最后一节给出了本研究的结论和未来改进的方向。

2 相关工作

电脑辅助诗歌创作这一领域已有不少前人的工作,中国台湾元智大学的罗凤珠教授^[5]提供了一个自动检查诗词格律的系统和一个帮助用户查找同韵字的字典。李良炎^[6]探讨了中文诗词风格评价技术。至于自动生成诗歌,网上广泛流传的“稻香老农”作诗机采用将辞藻按照语法直接填入诗词模板而成,生成的诗歌较为生硬晦涩。目前还没有看到较好的生成中文诗歌的结果。在中文对联方面,微软亚洲研究院的自然语言计算组研发的计算机自动对联系统^[7]首次采用了统计机器翻译用于对联生成。

其他语言的利用计算机生成诗歌研究工作始于1959年,Theo Lutz用计算机产生了第一首德文诗

歌^[8]。Mnuring在其论文中提出了一个完整的作诗机模型^[10]。另外,前人还设计了 RACTER 和 PROSE 等作诗系统^[8]。

总结前人对于诗歌生成的研究,可以将诗歌生成粗略地分为三类:基于模板的、进化的和基于范例的过程。就我们所知,目前还没有用统计的方法研究中文格律诗的工作发表。我们的工作可以说是这一领域的第一个尝试。

3 计算机绝句生成的总体流程

模拟人写绝句的过程,我们把计算机绝句生成的过程分为以下几个步骤:

1. 用户选择诗的形式(五言或七言),并通过输入关键词确定想要表达的内容;
2. 由计算机自动生成诗的首句;
3. 根据生成的首句并参考用户给定的信息,由计算机依次生成余下的诗句。

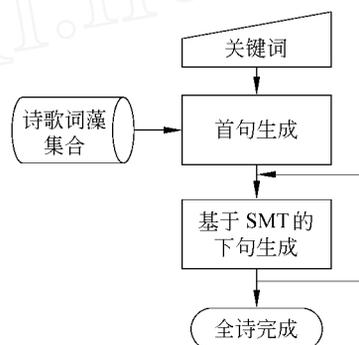


图1 绝句生成框架

用户首先选择待作绝句的内容可能涉及到的若干个语义类别,例如“时令类”、“游眺类”等。然后在各语义类别下选择一个关键词,比如分别选择“早春”、“踏青”等。这里所用到的语义类别以及每个类别下的关键词都是由《诗学含英》中的分类体系整理得到的。《诗学含英》中的分类体系共有40大类,1016个关键词,41218个词(其中不重复的34290个),其中的词汇长度从2到5不等。例如时令类是一个大类,里面有春色,早春等关键词。例如任意一个关键词为“春色”,书中将描写“春色”的诗歌中的一些辞藻整理归纳出来,得到与关键词“春色”相关

<http://www.poeming.com/web/index.htm>

清朝乾隆年间,山阴刘文蔚编辑之《诗学含英》根据《增广诗韵全璧》一书所附《诗学含英》建立,旨在提供初习作诗者应用辞藻典故之参考。

的辞藻,如山青、芳草等相关词汇。

当用户确定好诗的形式和三个关键词以后,我们利用一个相关词汇库和统计语言模型自动地生成若干首句候选。

有了绝句的第一句,要让电脑生成余下的三句,我们把绝句的上下相邻两句的关系类比于翻译问题中源语言和目标语言的对应关系,给定绝句上句,利用统计机器翻译的方法来“翻译”得出下一句。在生成第 N 句诗的时候,依据的信息是上一句诗,以及上溯到第一句的所有诗句,甚至包括用户输入的关键词。

我们提出的诗歌生成框架使得用户和系统之间的交互很方便。我们可以充分地利用用户的选择,改进系统的结果,最大限度地满足用户需求并产生质量良好的诗句。

4 首句生成

根据观察,五言绝句的诗句通常是以下七种节奏格式:(***)**代表一个三音节短语加上一个双音节短语,以此类推),*|****,****|*,**|**,**|**|*,**|**|**,**|**|*,**|**|*。而七言绝句的则主要是以下六种格式:**|**|****,**|**|**|**,**|**|**|**,**|**|**|**,**|**|**|*。假设所要生成的绝句的首句包含的词都是与用户选定的关键词相关的,那么只要我们将与这些关键词相关的词汇按照上述格式进行组配,就可以得到符合要求的绝句首句了。

按照这种方法来生成绝句首句,必须首先建立一个相关词汇数据库。本研究中,我们提出了利用《诗学含英》中的词汇分类体系来建立该数据库。对于500首格律诗的调查证明,95%的诗歌的第一句中出现的字词都可以在《诗学含英》中找到,因此用《诗学含英》作为诗歌词汇的来源是切实可行的。

利用根据《诗学含英》建立的相关词汇数据库,我们可以为用户选定的每个关键词生成一个相关词汇集合。这样,我们得到了若干个包含不同长度词汇的集合,然后把所有相关词汇依次从句子的第一个位置放置到句子的最后一个位置,生成一个词图。随后在词图上面采用 Forward-Viterbi-backward-A*搜索算法得到需要的首句。为了防止生成的诗句出现“孤平”等不合韵律的现象,我们加入了一个过滤过程,将不符合韵律要求的句子去掉。

举例而言,用户选定三个关键词:春日、郊行、访

友,要求生成五言绝句。我们的系统将会构建一个庞大的相关词汇表{“明媚”,“寻芳草”,“水村”,“旧话”,……},并且把相关词汇依次放置到句子的各个位置,形成的词图如表1。

表1 词图举例

字符1	字符2	字符3	字符4	字符5
明媚		寻芳草		
晴光		鱼	蝶飞	
花变新红				绽
...

接着我们可以采用 Forward-Viterbi-backward-A*算法得出 N 个候选首句,例如“晴光寻芳草”,“晴光鱼迎门”,“江山丽蝶飞”。

5 二、三、四句诗歌生成

5.1 基于统计机器翻译模型的诗句生成

之前的研究^[1]已经证明了统计机器翻译方法可以被用于对联下联的生成。对联可以被看作一种特殊的中文诗,且绝句中也频繁的包含类似对联的句对,例如我们统计了大量的绝句第一句和第二句,第三句和第四句,发现这些句对中约90%的句对存在着上下句结构一致的特点。同样地,绝句的第三句和第四句也有约90%的情况结构相同。所以我们希望利用同样的方法来生成绝句下句。不过,绝句下句的生成与机器翻译还是存在一些不同:源语言句子长度等于目标语的句子长度;跟对联一样,由于绝句的上下句子的字或者词一一对应,因此无需做词汇的对齐,也不存在词序的调整问题。所以,我们需要将统计机器翻译方法作适当的修改以用于绝句下句的生成。

同时,为了保持要生成的目标语句和上文的所有已有诗句的情境较为一致,保持绝句的完整性,使得绝句满足“起承转合”的结构要求。我们设计了一个互信息模型的特征函数用以补充。

具体而言,给定绝句的第 i 个句子 $F = \{f_1, f_2, \dots, f_n\}$ ($i = 1, 2, 3$) 和前 $i - 1$ 句诗歌,记作 $P = \{p_1, p_2, \dots, p_n * (i - 1)\}$,我们的目的是要找到一个第 $i + 1$ 个句子 $S = \{s_1, s_2, \dots, s_n\}$,使得概率 $p(S|F, P)$ 最大,其中 f_i , s_i 和 p_i 都是中文单字。根据文献[3]的结果,我们弃用了传统的噪声信道模

型,使用了一个更为普遍的对数线性模型:

$$S^* = \arg \max_S p(S | F, P) \\ = \arg \max_S \sum_{i=1}^M i \log h_i(S, F, P) \quad (1)$$

其中 M 是特征函数的总数, $h_i(s, f)$ 是不同的特征函数,具体如表 1 所示。在这个模型中,单字代替了词语,作为翻译基本单元来形成短语。这是因为中文诗歌语言精炼,绝大部分的词语都是由单字组成,因此引入中文分词反而会增加错误率。

本研究中,根据前人的工作^[1],我们采用了五个已经用到的特征函数(表 2 中的 h_1 到 h_5)。另外我们针对诗歌的生成的特殊性设计了第六个特征函数(表 2 中的 h_6)。在这里, $h_1 \sim h_4$ 依赖于 S 和 F , h_5 只由 S 决定而 h_6 由 S , F 和 P 共同决定。 h_6 直接代入公式,和前五个参数共同计算目标句子的分数。

为了应用基于短语的特征函数, S 和 F 分别被切分成 $\overline{s_1} \dots \overline{s_l}$ 以及 $\overline{f_1} \dots \overline{f_l}$,

表 2 对数线性模型的六个特征函数

$h_1(S, F) = \sum_{i=1}^l p(\overline{f_i} \overline{s_i})$	基于短语的翻译模型
$h_2(S, F) = \sum_{i=1}^l p(\overline{s_i} \overline{f_i})$	基于短语的反向翻译模型
$h_3(S, F) = \prod_{i=1}^n p_c(\overline{f_i} \overline{s_i})$	基于字的翻译概率
$h_4(S, F) = \prod_{i=1}^n p_c(\overline{s_i} \overline{f_i})$	基于字的反向翻译概率
$h_5(S, F) = p(S)$	语言模型
$h_6(S, F, P)$	互信息模型

注意到这里我们是以短语作为翻译模型的划分单位。这里的短语指的是在训练语料中出现过的任意字串,参考前人的研究,我们也将短语的最大长度设置为 4。

® 基于短语的翻译概率

对于诗歌上下句相同位置的任意两个短语 $\overline{s_i}$ 和 $\overline{f_i}$,我们可以认为 $\overline{s_i}$ 直接“翻译”为对应位置的 $\overline{f_i}$,因此翻译概率可以直接计算为它们在上下句对照语料中同时出现的频率:

$$p(\overline{f_i} | \overline{s_i}) = \frac{\text{count}(\overline{f_i} | \overline{s_i})}{\sum_{r=1}^m \text{count}(\overline{f_r} | \overline{s_i})} \quad (2)$$

其中 m 是短语 $\overline{s_i}$ 能够翻译为不同目标短语 $\overline{f_r}$ 的数目, $\text{count}(\overline{f_i}, \overline{s_i})$ 则是 $\overline{f_i}$ 和 $\overline{s_i}$ 同时出现在诗词上下句对应位置的次数。

® 基于短语的反向翻译概率

之前的研究表明,基于短语的反向翻译概率 $p(\overline{s_i} | \overline{f_i})$ (即目标语言短语 $\overline{f_i}$ 反向翻译为源语言短语 $\overline{s_i}$ 的概率)对于选择好的翻译候选具有补充作用^[4],因此我们在本研究中也把这一概率作为一个特征函数。其计算公式同公式(3),只是把源语言和目标语言的训练语料交换即可。

® 基于词汇的翻译概率

另外,我们需要较好地评价一对双语互译短语对之间的翻译质量^[2]。因此基于词汇的翻译概率必须用作一个特征函数,以评价短语对之间每个单词翻译为另外一个短语对的对应单词的概率。与基于短语的翻译概率的计算公式类似,词汇 s_j 翻译为对应词汇 f_j 的概率 $p_w(f_j | s_j)$ 可以按如下方式计算:

$$p_w(\overline{f_i} | \overline{s_i}) = \prod_{j=1}^{N_i} p(f_j | s_j) \quad (3)$$

其中 N_i 是 $\overline{f_i}$ 或者 $\overline{s_i}$ 中的单词数目, f_j 和 s_j 分别是 $\overline{f_i}$ 和 $\overline{s_i}$ 中的字,且 $p(f_j | s_j)$ 是字 s_j 翻译为 f_j 的概率。与短语翻译概率类似, $p(f_j | s_j)$ 还可以由单字的相关频率得出:

$$p(f_j | s_j) = \frac{\text{count}(f_i, s_i)}{\sum_{r=1}^m \text{count}(f_r, s_j)} \quad (4)$$

其中 m 是短语 $\overline{s_i}$ 能够翻译为不同目标短语 $\overline{f_r}$ 的数目, $\text{count}(f_i, s_i)$ 则是 $\overline{f_i}$ 和 $\overline{s_i}$ 同时出现在诗词上下句对应位置的次数。

® 基于词汇的反向翻译概率

与基于短语的翻译概率相仿,我们在翻译模型中也使用了反向的基于词汇的翻译概率 $p_w(s_j | f_j)$ 作为对 $p_w(f_j | s_j)$ 的补充。

® 语言模型

前四个特征函数主要是保证生成的下句与上句结构和语意的联系,而语言模型特征则是为了保证生成的下句符合诗歌的语言习惯。这里,我们使用了由 Katz 提出的回退权重训练得到的一个基于单字的三元模型^[13]。

® 互信息模型

在统计机器翻译和对联生成的工作中,目标语句只依赖于源语句。然而,在诗歌生成中,目标语句(即第 $i+1$ 个要生成的诗句)不仅仅依赖于源语句(即第 i 个诗句),同时也要受到前 $i-1$ 个诗句的影响。为了保持要生成的目标语句和上文的所有已有诗句的情境较为一致,我们设计了一个互信息模型的特征函数,计算方法如下:

$$h_6(S, F, P) = \sum_{x=1}^n \left[\sum_{y=1}^n MI(s_x, f_y) + \sum_{z=1}^{n*(i-1)} MI(s_x, p_z) \right] \quad (5)$$

任意两个字之间的互信息分数的计算公式如下:

$$MI(c_1, c_2) = -\log \frac{p(c_1, c_2)}{p(c_1)p(c_2)} \quad (6)$$

其中 c_1 和 c_2 是两个单字, $p(c_1, c_2)$, $p(c_1)$ 和 $p(c_2)$ 分别代表 c_1 和 c_2 同时出现的概率, c_1 出现在诗歌中的概率和 c_2 出现在诗歌中的概率。这些参数可以由对训练诗歌语料采用 MLE 方法得到。

5.2 模型的训练

® 训练数据

利用互联网我们获得了《全唐诗》、《全宋诗》、《全宋词》、《明诗》、《全清诗》和《全台诗》,所有语料共计 30 多万首古典诗歌和 350 多万个单句,所有的诗歌都用作互信息模型的训练语料。同时,从中我们提取了 120 多万上下句对照句对用以训练翻译模型。从一首绝句中可以提取到三对上下句对照句对(第一句对第二句,第二句对第三句,第三句对第四句)。由于纯粹的绝句数量有限,所以为了获得更多的训练数据,我们也从其他诗歌里提取了对照句对(例如宋词中前后相同长度的句子),但是要求上下句字数一致。

® 翻译概率插值

即依照所要生成的诗句在整首绝句中的具体位置采用不同的翻译概率表组合。绝句的起承转合四项任务,一般分由第一句、第二句、第三句和第四句来分别完成。所以,不同位置的诗句,理应采用不同权重的翻译概率表。因此,我们搜集了所有绝句的三种翻译组合(第一句对第二句,第二句对第三句,第三句对第四句),分别训练得到三个不同的翻译概率模型,并对其进行插值。

本研究中,我们将绝句对应位置的上下句对照训练语料得到的翻译概率赋以较大权重,而将从其他句对训练得到的翻译概率赋以较低的权重。举例而言,在根据第一句翻译第二句时,所有从绝句第一句和第二句训练得到的翻译概率模型都将被赋以较大权重,而所有诗句句对训练出的翻译概率模型将被赋以较小权重。具体实现时,我们确定较大的权重为 0.8,较小的权重为 0.2。每个位置的解码使用的翻译概率表即为不同位置的翻译概率表的插值结果。

® 语言模型间的插值

我们的语言模型训练数据分成两个部分:来自古典诗歌的 350 多万个单句,互联网上获取的约 1 200 万行的古文语料。利用困惑度的公式,我们对训练数据计算出了 Katz 提出的回退权重,根据最大似然估计分别训练了一个基于单字的三元模型:古典诗歌语言模型 $p_1(s)$ 和古文语言模型 $p_2(s)$,随后进行线性加权得到最终的语言模型如下

$$p(s) = 0.8p_1(s) + 0.2p_2(s) \quad (7)$$

5.3 解码

为了适应绝句生成的某些特殊要求,我们对传统的统计机器翻译解码器^[2]进行了一些修改,首先不允许词语调序,这使得我们的解码器是单调的。另外由于诗歌的句子长度一般都小于机器翻译中的句子长度,因此我们的解码器效率比典型的机器翻译解码器的效率更高。此外,我们还进行了关于韵律方面的修改:

首先,格律诗的首句不允许“孤平”的出现,我们直接在选取前 N 个候选句子的时候,就去掉了可能产生“孤平”的词语。

其次,根据绝句的押韵要求,第四句的最后一个字(称为“韵脚”)的韵必须和第二句韵脚的韵相合,两个韵脚的平仄也需要一致。

为了满足此押韵要求,在解码第四句时,我们需要对最后一个字或者短语作特殊处理。首先,我们需要根据第二句韵脚字的韵和平仄,删掉翻译候选中不符合要求的字或者短语。其次,为了使最后一个字的位置能有足够多的候选以保证生成结果的多样性,我们在过滤后留下的候选里新插入了一些满足韵律要求的候选字。例如,如果我们知道某诗第二句的韵脚是“眠”,因此按照古典诗词韵书《平水韵》中所载的汉字古韵选取与“眠”字有相同韵脚的所有字,同时去除掉平仄不符合的字。然后我们在翻译概率表中动态插入这些提取的字。使得我们的翻译出来的句子满足押韵要求。

6 实验

绝句的评测是一个很关键的问题。之前的自动

包括南朝萧统的《文选》、《唐宋八大家散文选》等。

南宋平水刘渊,将同用的韵合并,成 107 韵,后人渐为 106 韵,被称为平水韵,一般叫“诗韵”。

对联研究^[1]采用的是基于 BLEU^[14] 的评测方法。具体来讲,就是给定若干个上联,人工为每个上联写出多个下联。然后将这些上联作为输入,让系统为每个上联生成一个下联,最后计算系统生成的这些下联与人写出的下联的 n-gram 重合度。

然而,我们没有采用这种自动评测系统,原因是该评测系统需要人工给出足够多的“标准答案”,而对于绝句生成而言,给定相同的关键词,“标准答案”的发散性太大。所以,在本研究中,我们采用的是人工评测方法。我们首先建立了一套翔实完整的人工评测标准,然后采用 blind test 让不同的人给系统产生的诗歌进行打分,取平均分用以评价生成结果的好坏程度。这样的评价方法在给定的实验数据较大(20~40组)的情况下尽力消除了主观性。

为了较好地评测我们的方法产生的诗歌水平,我们共进行了三项针对不同内容的实验。

® 实验一:首句评测。即给定三个关键词,对自动生成的绝句首句进行评测。

® 实验二:下句评测。即给定绝句上句,对自动生成的绝句下句进行评测。

® 实验三:全诗评测。即加入人工交互后对全诗水平的评测。

6.1 首句评测

我们首先随机选定一组关键词,让系统生成 20 个首句,然后人工选出一句最好的首句,根据诗句语言,与关键词的联系程度,以及是否合适作为首句这三个标准,加权评分。各权重和分数如表 3 所示。

表 3 首句评测标准

评测标准	权重	不接受	可接受	良好
诗句语言	2/5	0	50	100
与关键词的联系程度	2/5	0	50	100
作为首句是否合适	1/5	0	50	100

在本实验里,我们一共随机选取了 40 组关键词,其中 20 组五言首句,20 组七言首句,得到的平均分见表 4。

表 4 首句评测结果

五言	七言
69.5	78.5

可以看出,七言诗句的生成效果要好于五言诗句,这是因为我们在评测首句生成的标准里面,强调

了与用户所给关键词的符合程度,这样七言诗句的涵盖内容更广泛,对关键词的命中率也提高了。对于得分较低的关键词组,我们进一步分析发现其主要原因是我们在实验时采用了随机选取关键词的办法,选出的关键词内在联系不紧密,导致了生成的首句准确率不够高。事实上,用户通常会选择有一定联系的关键词,这样我们的系统就可以发挥更好的效果。

6.2 下句评测

我们从首句实验中依次选出 10 个最佳的首句,并依次生成第二句,然后选出最佳的第二句,生成第三句……以此类推可以得到若干组最佳下句,并交给评测员打分。评测标准见表 5。

表 5 下句评测标准

评测标准	权重	不接受	可接受	良好
诗句的语言	1/2	0	50	100
和上句的相关性	1/2	0	50	100

由此我们可以得到每个下句的平均分,随后把所有符合要求的诗句分成良好和可接受这两个等级,标准见表 6。

表 6 下句的不同等级

等级	良好	可接受
分数	= 100	> = 50

若前 N 个候选诗句中,存在着至少一个良好或可接受等级的诗句,那么该测试例子符合要求。最后我们分别计算符合这样要求的测试例子的比例,得到结果见表 7。

表 7 五言诗句的测试结果

评价项目/等级	良好				可接受			
	前 1	前 5	前 10	前 20	前 1	前 5	前 10	前 20
第一句生成第二句	17%	37%	50%	60%	63%	87%	90%	97%
第二句生成第三句	20%	43%	53%	63%	63%	93%	97%	97%
第三句生成第四句	17%	37%	47%	63%	33%	80%	87%	97%

可见,有 50%左右的实验样例,前 1 句中存在着一个可接受的诗句。而前 5,前 10 和前 20 中,至少存在一个可接受诗句的实验样例比例依次增大,最高可达 80%。

表8 七言诗句的测试结果

评价项目/等级	良好				可接受			
	前1	前5	前10	前20	前1	前5	前10	前20
第一句生成第二句	0	40%	53%	60%	67%	80%	87%	97%
第二句生成第三句	25%	44%	50%	63%	50%	63%	75%	94%
第三句生成第四句	11%	39%	44%	56%	44%	67%	78%	83%

实验中也能看出一些不足之处,从总体上说,A等级的诗句的比例并不高,也就是说我们的系统能作出符合要求和格律正确的诗句,然而,系统较少能作出特别精彩或者感情丰富的诗句。未来我们可以考虑对古典诗句进行词语标注甚至情感标注,从而提高诗句的语义水平,丰富诗句的情感。

6.3 全诗评测

为了评价我们的方法产生的完整诗歌的水平,并与已有的其他作诗机进行比较,本实验中,我们共选定了20组关键词,生成了20组绝句(五言绝句10组,七言绝句10组),并用同样的关键词下稻香作诗机生成的诗歌进行比较。其中,我们采用了自动生成和人工交互生成两种手段:第一种是人工交互生成,即给定关键词,当系统输出了每一句的候选之后,人工交互选出最好的一句作为下一句的输入。第二种是自动生成,即在交互的每一步,由计算机自动生成唯一的最优解,并作为下一句的输入。

最后,我们请评测员对20组生成的诗歌,按照表9给出的评测标准打分。每组诗歌均包括稻香作诗机的结果,我们的自动作诗机的结果和我们的人工交互方法的结果(见表10)。

表9 全诗评测标准

评测标准	权重	不接受	可接受	较好
语言	5/15	0	50	100
韵律	5/15	0	50	100
是否符合题意	3/15	0	50	100
结构	3/15	0	50	100
意境	1/15	0	50	100

每首诗歌的分数得到之后,我们可以计算出三种方法的平均分数,如下表所示。

表10 全诗评测结果

	稻香作诗机	自动作诗机	人工交互作诗机
Avg. score	54.83	68.34	77.83

从表10可以看到,我们的两种方法得到的结果都要优于稻香作诗机。在交互模式下,生成诗歌的平均分数为77.83分,意味着大部分的诗歌都是可接受的,且近一半的诗作可以被认为是较好的诗作。考虑到作诗的难度,这个分数已经不错了。

表11给出关键词为“豪放,诗,父子”以及关键词为“秋日书怀,秋望,月影”的例子,可以看出,尽管语言和结构的差距不是特别明显,但是我们的方法产生的诗歌比之稻香作诗机要切题得多,语言也更加优美古朴。例如关于“豪放诗歌”,我们的作诗机用“六朝风月”这个典故来形容诗歌的内容,用“狂吟醉舞”来形容诗人作诗的过程,用“大梦醒来烟柳老”来描述完成诗歌创作后诗人的心情,有一定的艺术性。

表11 生成的诗歌

关键词“豪放,诗,父子”		关键词“秋日书怀,秋望,月影”	
稻香作诗机	自动作诗机	稻香作诗机	自动作诗机
定被吟成年渐长 今春史已献名王 炉香市里西施面 吃酒愁人卧草堂	六朝风月一时清 万里江山千日好 狂吟醉舞墨花香 大梦醒来烟柳老	权臣自此尽成愁 有使今来往扁舟 问雪乌台游柱史 寒通柳径佛宫秋	一雁秋色三更雨 千山白云万里风 桂影蝉声倦登楼 花光月色愁入梦

7 总结

本文独创性地结合了统计机器翻译的模型和基于传统词藻分类的生成模型来自动生成中文格律诗。在用户给定几个关键词后,我们的系统首先利用已有的

分好类的词藻来生成一个语言模型分数最高的诗句。随后我们应用统计机器翻译的技术,结合邻近的诗句和全诗的语境从诗歌的上句翻译生成诗歌的其他句子。根据严格的人工评测的结果,我们的方法能够生成比较合理的诗歌,并且比已有的一个中文诗歌生成器准确率高。

未来我们的工作包括以下几个方面:首先,采用首句模板生成和统计机器翻译模型来进行绝句的生成,并不是十全十美的。我们已经考虑了绝句作为一个整体的影响,并加入了互信息模型,然而未来可以考虑采用别的更好的办法来加强整个绝句的整体性和完整性。并对韵律要求更加严格。

在对特征函数进行评价的过程中,我们考虑采用固定其他特征函数,对某一个特征函数的作用进行实验研究,得到定量数据,这样能够使得若干特征函数的权重更加科学合理。

还有,古典绝句的格律要求十分严格,现代人写绝句已经不拘泥于这样严格的格律要求了,同时古音(特别是平仄)到今天已经发生了改变。因此我们的系统只考虑了押韵要求,没有考虑平仄要求。未来我们可能根据《平水韵》将古汉语的平仄要求加入作诗机,生成韵律更为精准的诗歌。

另外,我们还考虑对产生的诗歌和已有的古典名诗混合进行“图灵测试”,通过测试可以判断,一个懂得一定诗歌知识但不是诗歌专家的读者能否正确地区分诗人写出的诗歌和电脑写出的诗歌。

参考文献

- [1] Long Jiang, Ming Zhou. Generating Chinese Couplets using a Statistical MT Approach [C]// The 22nd International Conference on Computational Linguistics, Manchester, England, August 2008.
- [2] Philipp Koehn, Franz Josef Och, and Daniel Marcu. Statistical Phrase-Based Translation [C]// HL T/ NAACL 2003.
- [3] Franz Josef Och. Minimum Error Rate Training for Statistical Machine Translation [C]// ACL 2003: Proc. of the 41st Annual Meeting of the Association for Computational Linguistics, Japan, Sapporo, July 2003.
- [4] Franz Josef Och, Hermann Ney. Discriminative Training and Maximum Entropy Models for Statistical Machine Translation [C]// ACL 2002: Proc. of the 40th Annual Meeting of the Association for Computational Linguistics, Philadelphia, PA, July 2002: 295-302.
- [5] 罗凤珠,李元萍,曹伟政. 中国古代诗词格律自动检索与教学系统 [J]. 中文信息学报, 1999, 13(1): 35-42
- [6] 李良炎,何中市,易勇. 基于词连接的中文诗词风格评价技术 [J]. 中文信息学报, 2005, 19(6): 98-104.
- [7] 苏劲松,周昌乐,李翼鸿. 基于统计抽词和格律的全宋词切分语料库建立[J]. 中文信息学报, 2007, 21(2): 52-57.
- [8] Charles O. Hartman. Virtual Muse: Experiments in Computer Poetry[M]. Wesleyan University Press, 1996.
- [9] Naoko Tosa, Hideto Obara and Michihiko Minoh. Hitch Haiku: An Interactive Supporting System for Composing Haiku Poem[C]// ICEC 2008: 209-216.
- [10] H. Manurung, G. Ritchie and H. Thompson. Towards a computational model of poetry generation[C]// Proc. of the AISB-00 Symposium on Creative and Cultural Aspects of AI, 2001.
- [11] Franz Josef Och, Nicola Ueffing, Hermann Ney. An Efficient A* Search Algorithm for Statistical Machine Translation [C]// Data-Driven Machine Translation Workshop, Toulouse, France, July 2001: 55-62.
- [12] Philipp Koehn. Pharaoh: a beam search decoder for phrase-based statistical machine translation models[C]// Proceedings of the Sixth Conference of the Association for Machine Translation in the Americas, 2004, pp. 115-124.
- [13] Andreas Stolcke. SRILM—An Extensible Language Modeling Toolkit [C]// Proc. of Intl. Conf. on Spoken Language Processing, 2002, 2: 901-904.
- [14] Kishore Papineni, Salim Roukos, Todd Ward, Wei-Jing Zhu. BLEU: a Method for automatic evaluation of machine translation[C]// Proc. of the 40th Meeting of the Association for Computational Linguistics, 2002.