**基于LSTM模型的现代文古诗生成技术**

# 一 背景介绍

读史书使人明智，读诗书使人灵秀，诗词是中国千百年来的文化积淀和民族瑰宝，是语言的精华，是智慧的结晶，是思想的花朵，是人类最纯粹的精神家园。但随着语言的发展，当代人对诗词的了解越来越少，作诗的能力也远远比不上古人。

幸运的是，随着大数据时代的到来，人工智能飞速发展，通过计算机古诗词的自动转换成为可能。例如，可以通给出一个目标主题，机器可以自动生成一首符合该主题的古诗。让机器做诗可以减少人类在写诗中寻词、造句上的困难，使诗人关注诗的主题、意境等更有价值的东西；同时，机器做的诗不受人类传统思维的束缚，创新性极强。这些创新性虽然并不完全合理，但却给诗人丰富的提示，进而激发诗人更多创作灵感，写出更具有奇思妙想的创新型作品。我们相信，古诗的自动生成可以极大激人们对古典文学的兴趣，对我国传统艺术的发展与继承有着重要的作用。

然而，当前依主题指定方式为基础的自动做诗方法具有很大的局限性。一方面，这种方式对主题词的选择要求较高，只有主题词选的合理，生成的古诗才更合理。这对很多普通用户形成很大障碍，我们希望用户只用现代汉语描述出要生成古诗要表达的意境和内容，机器将这一描述浓缩成古诗，因而可极大提高古诗生成系统的适用性。另一方面，依当前主题词指定的方式，诗人只能在大泛围内给出做诗的内容，却不能精确控制做诗的流程。事实上，古诗不仅是词语的合规批凑，更重要的是诗人感情通过写景、叙事过程的递进式升华。当前以主题限定方式生成的古诗，不具有思维的连惯性和活跃性，无法形成合理的表义逻辑，无法形成完整递进的主题表达，因而生成的诗很多只是词语的堆砌，缺乏情感和逻辑性，应用有限。

本发明提出一种新颖的古诗生成方式和相关算法，可以将一段现代汉语浓缩转写为古诗词；同时，允许现代文按句指定每句古诗的内容，这样现代文中表现出的起承转合的逻辑流程可以反映在生成的古诗中，为古诗生成带来灵魂。

# 问题描述

与传统主题词指定的做诗方法相比，现代文古诗生成要困难的多。一个显著的困难是，现代文的长度一般远大于古诗长度，如果要将现代文转化为古诗，需要提取出其中最有价值的语义信息，并将这些信息用精练的古代汉语表达出来。比如：“现在正是早春二月，春风吹拂，春草萌生，过不了多久，就会绿透江南岸边”可以对应诗词“春风又绿江南岸”，白话文中的“现在正是”，“过不了多久”，“就会”这些词都对表义没有太多帮助。我们需要提取出如“绿”，“江南”，“岸边”等表意比较明确的字，形成句子的意义。如何提取出句义，对合理的古诗生成具有重要意义。

另一方面，一句现代汉语可能有多种转写形式， 如何自然地生成多种表达方式的古诗，也是需要解决的问题。

本发明提出了一种现代文诗词生成系统。该系统将神经网络与规则相结合，通过神经网络对现代汉语进行语义理解，再依规则系统生成符合语义的古诗。为保证生成古诗的多样性，我们采用两种引入随机性的办法，一是在对输入的现代文进行一些随机扰动，二是对文体规则进行更改。前者引入表达多样性，后者引入文体多样性（如五言诗、七言诗、多种韵律格式等）。

这一发明的优势是：

1. 通过计算机自动将现代文转化成古诗，更符合大众使用习惯
2. 通过将现代文中的逻辑性表达成古诗表达上的逻辑性，摆脱词语堆砌，形成合规且具有表义能力的古诗。
3. 通过加入随机性可以写出具有多种表达、多种文体的丰富多彩的古诗。

# 三 发明要点

## 3.1 通过神经网络的古诗生成模型

本发明基于LSTM的编码-解码网络构造现代汉语-古诗转换模型，如图一所示。首先，图一的下半部分，系统将用户的现代文输入句子经过一个双向LSTM网络编码成语义向量，其中每个句子表达成一个语义向量，形成一个语义向量组。这些语义向量作为用户意图的编码。在生成过程中（图一的上部），一个LSTM网络不断循环运行，生成文本的中的每一个字。在生成每**一个字**的时候，需要将**该句**所对应的语义向量作为参考输入，使得该句生成与用户要求的表义内容相符合。

注意，在生成过程中，我们加入断句、押韵、平仄等文体规则，保证了生成的“字串”既能最大程度地符合用户意图的同时，遵循强制的诗词规则。

h1

x1

x1

h1

h2

x2

x2

h2

h3

x3

x3

h3

h4

x4

x4

h4

h5

x5

x5

h5

h6

x6

x6

h6

at，1

at，2

at，3

at，4

at，5

at，6

st-1

st

…

…

yt-1

yt

文体规则

 **图一 基于RNN和LSTM的生成模型**

## 3.2 一致性保证

在上面的生成过程中，每句现代文生成一个语义向量，用以生成一句古诗，因此这句诗都会围绕这句的主题；同时，后一句的生成结果依赖前一句生成的词，因此保证整首诗的连惯性。这意味着我们的生成方法有能力生成一首前后连惯，同时又符合现代文逻辑线索的古诗，保证一致性。

## 3.3 多格式诗词生成

3.1中的模型会依赖一个文体规则约束，生成的古诗必须符合该文本要求。通过更改这一文体规则，可以生成多种格式的古诗词。比如，在七言律师中，这一文体规则要求每生成七个字后，必须生成一个结束符；同时，每句第七个字必须符合统一的韵律，每一句中的每个字需要符合一定的平仄要求。变换这一文体规则，可以得到不同韵律，不同体裁的古诗，基至可根据不同词牌（如虞美人、浪淘沙）设计文体规则，生成各种词牌的宋词。

## 3.4 诗词多样性生成

我们可以通过引入随机性，生成多种表达的不同古诗。比如，我们可以在输入的白话文中随机加入一些相近词，使生成的古诗具有多样性。这些随机性可通过“增加”、“删除”、“修改”等操作来实现。如“问题描述”中提到的例子：“现在正是早春二月，春风吹拂，春草萌生，过不了多久，就会绿透江南岸边”，可以通过删除“江南岸边”或者加上“鸟儿”，得到“草长莺飞二月天”。加入少量的随机化信息可以增加生成的诗词的多样性，且不会使最终生成的诗词离主题偏离过远。

# 四 方案优势

* 本发明可使普通使用者通过输入现代文生成诗词，比传统主题生成方法更便捷。
* 本发明可对古诗生成过程进行细致控制，使生成古诗具有逻辑线索。
* 本发明可通过各种随机性生成多种表达和样式的古诗，更加丰富多彩。