**基于LSTM的文本生成**

郭润堂

guoruntang@buaa.edu.cn

**Abstract**

LSTM 是一种能够处理长序列信息的循环神经网络，它具有记忆和忘记的机制，适合于文本生成这样的自然语言处理任务。本次作业的文本生成任务是使用LSTM算法根据给定的上下文，自动生成合乎语法和逻辑的文本。本作业采用了 10 本预处理过的武侠小说作为数据集，利用 word2vec 对文本进行编码和解码，然后使用LSTM算法进行训练，在给定一段话的基础上，预测下一段话，成功生成了一些具有武侠小说特色的文本。

**Introduction**

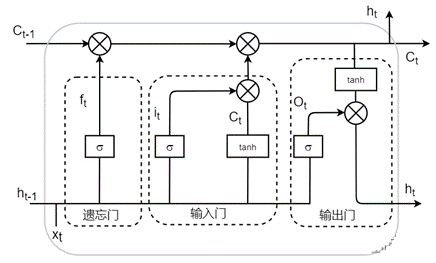


图 1 LSTM

文本生成是自然语言处理中的一个重要任务，它的目标是利用深度学习等技术，自动地生成可理解的自然语言文本。文本生成有多种应用场景，例如机器新闻写作、聊天机器人、数据到文本、图像到文本等。文本生成的方法和模型也很多，主要可以分为基于规则、基于模板和基于神经网络的三类。其中，基于神经网络的方法是目前最主流和最有效的，它通常采用编码器-解码器（encoder-decoder）的框架，利用递归神经网络（RNN）、长短期记忆网络（LSTM）、变分自编码器（VAE）、Transformer等模型来学习输入和输出之间的映射关系。文本生成的评价方法也有很多，例如使用BLEU、ROUGE等指标来衡量生成文本与参考文本的相似度，或者使用人工评价来判断生成文本的质量、流畅度、多样性等方面。文本生成还面临着一些挑战和难点，例如如何提高生成文本的语义连贯性和一致性，如何控制生成文本的句式和情感表达，如何利用知识库等外部信息来增强生成文本的内容等。

[LSTM算法是一种循环神经网络（RNN）的变体，它可以有效地解决RNN在处理长序列时遇到的梯度消失或爆炸的问题。LSTM算法的特点是引入了三个门结构（输入门、遗忘门和输出门），以及一个细胞状态，来控制信息的流动和记忆。LSTM算法可以学习长期依赖关系，捕捉序列中的时间特征和语义特征](https://zhuanlan.zhihu.com/p/366525011" \t "https://edgeservices.bing.com/edgesvc/_blank)。

[LSTM算法在自然语言处理文本生成领域中的作用是可以作为编码器-解码器（encoder-decoder）框架的基础单元，将输入文本编码为一个隐向量，然后将隐向量解码为输出文本。LSTM算法可以从字符级或单词级生成文本，根据已经观测到的字符或单词，预测下一个字符或单词出现的概率，从而实现文本生成2](https://zhuanlan.zhihu.com/p/188446640" \t "https://edgeservices.bing.com/edgesvc/_blank)[。LSTM算法也可以与变分自编码器（VAE）或注意力机制（attention）等技术结合，进一步提高文本生成的质量、多样性和可控性](https://blog.csdn.net/fgg1234567890/article/details/113532901" \t "https://edgeservices.bing.com/edgesvc/_blank)。

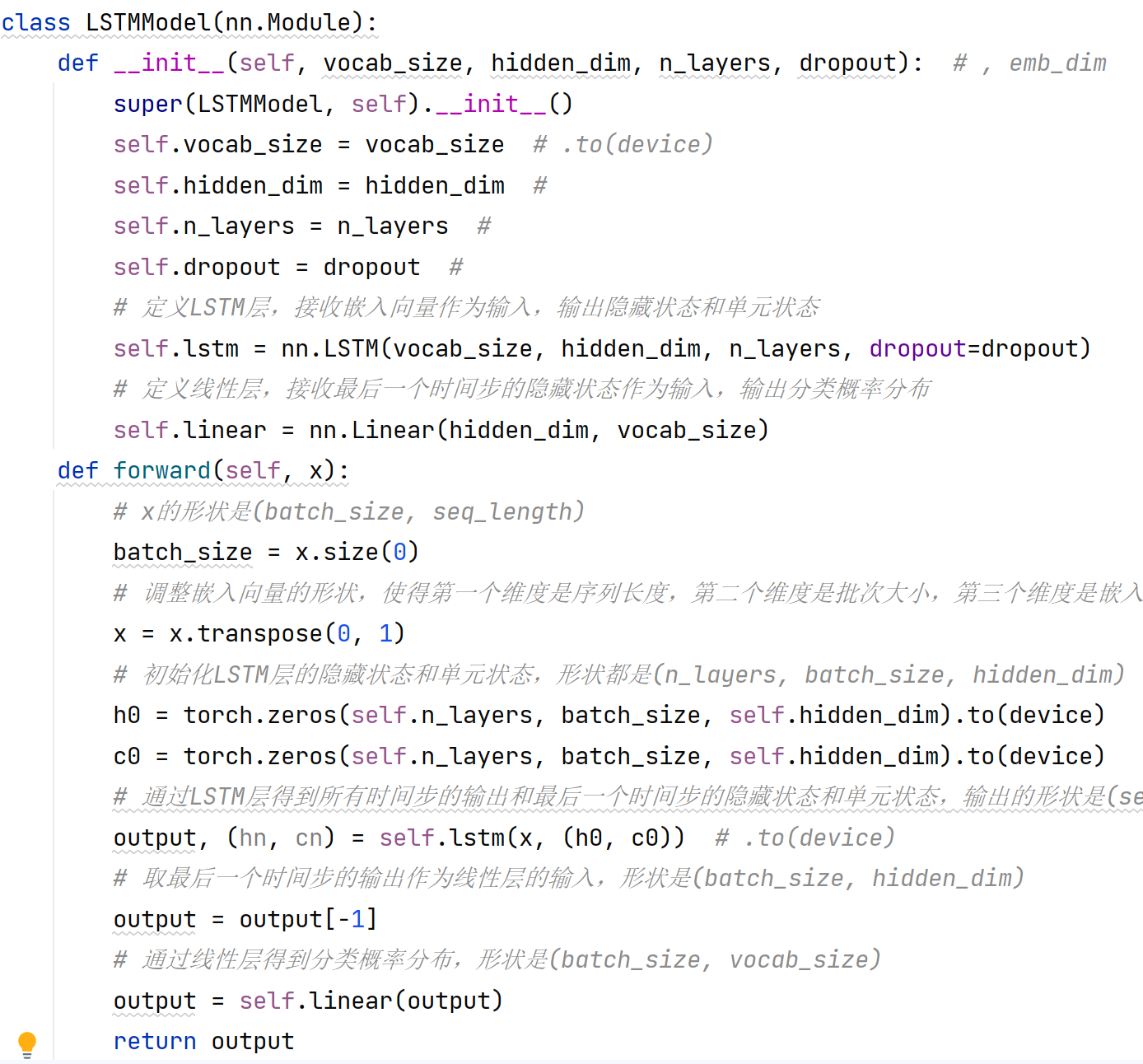
本次代码使用了Gensim，一个简单搞笑的自然语言处理库中的Word2Vec算法，将经过预处理后的文字进行向量编码作为LSTM算法的输入，通过对LSTM算法输出的向量与word2vec模型中的向量进行比较，选取相似度较高的字作为输出进行解码。

**Methodology**

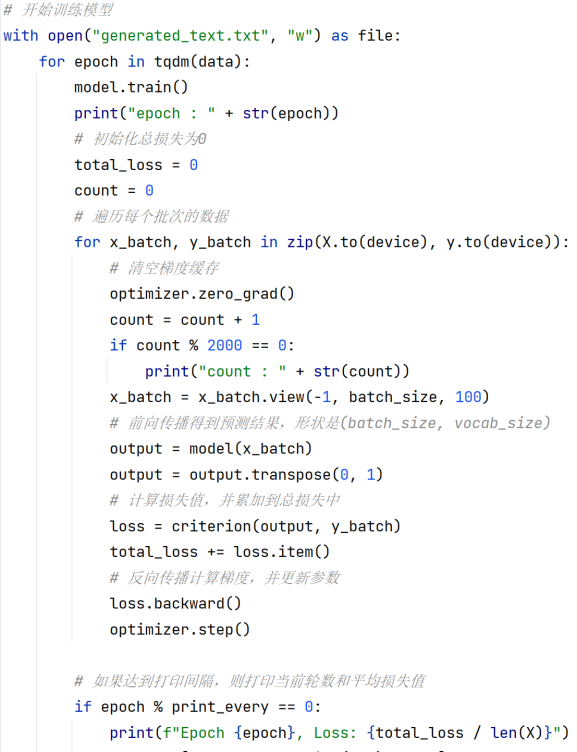
首先对文本进行预处理然后通过gensim的word2vec将文本中文字转换为256维的向量进行表示。



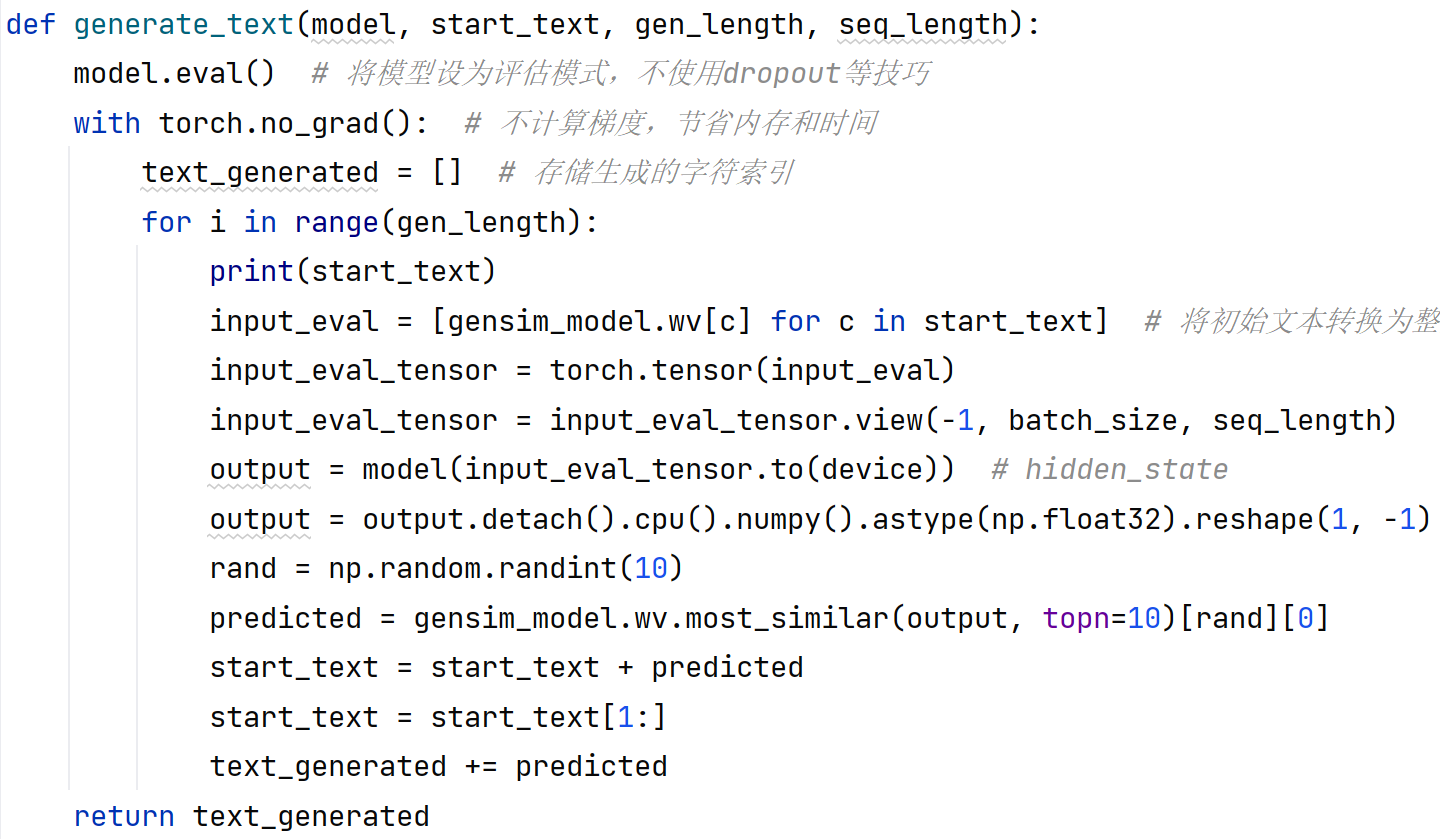
然后对LSTM网络进行定义



将数据库文本转为向量后，按照每20个一组的顺序依次输入输入网络，对下一时刻的文字向量进行预测，并与真实文字的向量进行比较。损失函数使用mse函数。对模型进行训练：



从训练集中按照顺序随机抽取20个字符，对后50个字进行预测



结果如下：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 输入文本  Epoch | 金兵到来 | 段誉伸出右手小指 | 他立刻闭眼，心中砰砰乱跳，过了片刻，才慢慢睁眼，只见所处之地是座圆形右室，光亮从左边透来，但朦朦胧胧地不似天光。 | 杨铁心正自力挣双手，急切无法脱缚，突见义兄受伤倒地，心中急痛之下，不知从哪里忽然生出来一股巨力，大喝一声，绳索绷断，挥拳打倒一名士兵，抢过一柄长矛，展开了杨家枪法，这时候一夫拼命，万夫莫挡。 |
| 10 | perplexity 24.999548  ，不知如何是好。他们是个大大的不解，只是他们的事，不知是否能够，这时候还不能再见到他。他们是个大燕子 | perplexity 24.999548  将他手指一捏，便即取过了他的手掌，将他手指在手掌上一拍，说道你这小子不是我的，我也不能再打你一拳。 | perplexity 25.218992  这一日来到一处大穴，一个个都是个女子，心中一喜，又是一阵凄凉，随即想起了这小子，不知是否能将她一个女 | perplexity 25.218992  这一下变招，一招一式，便已被他震断。那老者左手抓住他的右手，右手一探，已抓住了他的左手，右手一探，已 |
| 20 | perplexity 22.649045  ，不知如何是好。他们是个大大的好人，那是甚么人？那人道你们这么说，我们是个大大的，我们都是一样。那老 | perplexity 22.649045  ，将他手指一捏，便即解开他的穴道。这一下变化不守，只得将他手中的铁杖打得脑浆迸裂而死。段誉心想这小子 | perplexity 22.861043  那宫女道这是我们的大事，我们也不能跟你说。段誉道我们是在这里，我们也不能跟你说。阿紫道我不是你的，我 | perplexity 22.861043  这一下变招，只见他左手抓住了他的右手，右手一挥，将他身子一晃，便即毙命。张无忌一惊，心想这人武功虽高 |
| 30 | perplexity 21.682239  ，不知如何是好。他们是个大大的名号，但他们是个大大的好人，是不是？那老者道我们是个大大的名号，是不是 | perplexity 21.682239  ，向他胸口刺去。段誉心中一惊，只见他左手一伸，将他手腕一震，登时摔倒。段誉心中一惊，只见他双手抓住了 | perplexity 21.811128  他一惊之下，心中大喜，心想这一下变故，不知是否能将他们抛在地下，只怕他们不敢再行出去，只怕这里有人来 | perplexity 21.811128  这一下变招，便即摔倒，但见他双手已被缚住，双手一挥，将他身子一挡，摔在地下。那老者见他身形甚高，身形 |
| 40 | perplexity 21.066218  ，不知如何是好。他们是个大大的好人，那是不是一个人，不知是否能够抵挡？那宫女道这是我们的了。那宫女道 | perplexity 21.066218  ，左手食指点到，左手食指在他肩头一捺，露出手臂，右手食指在他肩头一捺，右手一伸，已抓住了他左腕， | perplexity 21.234074  那宫女道这位是大理国镇南王，不知是谁？段誉道我们是大理国的公主，是以来到大理，到了南方的大事，不能再 | perplexity 21.234074  张无忌一惊，只见一个身穿青布长袍的老者身子一晃，已将他手中长矛挡格，手一掌，将他手中长矛挡开。张无忌 |
| 50 | perplexity 20.645848  ，不能再见到他的。他们是个大大的好人，那是不能再见了。那老者道我们是个大大的好人，是不是？那老者道我 | perplexity 20.645848  ，向段誉胸口刺去。段誉心中一动，只道他是个大大的好人，只是他这一掌力道大得狠劲，不能再使劲，那是非同 | perplexity 20.803549  那女子一惊，心想这人是个女子，不知是谁？那人道我们是个大大的小子，那也不是好人，不知是否已死，我们也 | perplexity 20.803549  但这一招，那是再也不敢硬接，但他身上的劲力已然不及，只是一招一式，便即毙命。这一招武功高强之人，自己 |
| 60 | perplexity 20.337113  ，不能再见到他的手下。那人道你们这么一个人，你们都是我的。那老者道我们是个大大的名号，是不是？那人道 | perplexity 20.337113  ，一招一式，一拳打出，将他身子一撞，将他身子一撞，将他身子踢飞，左手抓住了他左手，右手抓住了他左手， | perplexity 20.477065  那宫女道我们在这里，我们在这里等候。那人道我们是大理国镇南王世子，不知是甚么路道了。那宫女道我们是大 | perplexity 20.477065  这一招，一招一式，便即收回，这一招，一招一式，便即收回，右手一招，右手剑交胸，右手剑尖已刺到他胸口， |
| 70 | perplexity 20.202380  ，不能再用兵刃，那是不能再打一个，但他们是个大大的好人，不知是否能够胜他？那老者道我们是个大大的，我 | perplexity 20.202380  ，左手一探，抓住了他的左腕。段誉双手一拍，一掌击在他胸口膻中穴，右手一掌拍出，右掌向段正淳胸口直 | perplexity 20.272065  他心中一喜，心想这小子是个女子，不知是否能有这等高手？那老者道我们是个大大的大哥儿，你们要去找他们报 | perplexity 20.272065  他心中一惊，暗叫不好，只见他双手已抓住了他的手腕，左手抓住了他右手，右手抓住了他的右手，右手一招，右 |
| 80 | perplexity 19.947092  ，不知如何是好。他们在这里，不知如何是好，不知如何是好，只是他们在这里，不知是否能够报答，那是不能再 | perplexity 19.947092  ，一招手法使得凶狠，心下暗暗惊异，只见他左手一伸，已抓住了他左腕，右手一扬，一枝袖箭向他射去。段誉心 | perplexity 20.038034  他心中一喜，忙道我不是我的，我不知道你是谁？那女子道我不是你的，我不是你的。那女子道我是你的，我不是 | perplexity 20.038034  但见他身形一晃，已然不及，左手抓住了他右肩，右手抓住了他的右手，右手食指已抓住了他的右眼。那人一声长 |
| 90 | perplexity 19.757391  ，不知如何是好。他们在这里，不能再见到他们。那老者道你们在这里，我们都是好的。那老者道你们在这里，我 | perplexity 19.757391  ，在他手腕上一点，左手一剑已将他身子踢中，左手一剑，向他胸口刺去。段誉心中一凛，暗想这小子的内力，我 | perplexity 19.899950  那农女道你们这些人都是武林中的大事，我们也不能跟你说了。那老者道我们这个人，你们要杀我，我们也不能跟 | perplexity 19.899950  这一下变招，只是一招双雕，一招双雕，双足在地下一撑，跃起，双足在地下一撑，跃起，左手抓住了他右手，右 |
| 100 | perplexity 19.614549  ，不知如何是好。他们在这里，我们在这里陪着我们。那老者道我们是个大大的好人，是不是？那人道你们是谁？ | perplexity 19.614549  ，在他手腕上一点，手指已被他抓住，这一下却是一拳打在段誉身上。段誉心想这小子武功高强，不能再跟我比武 | perplexity 19.756388  那宫女道我们是在缥缈峰上的大事，你们要去杀害慕容复，不能再去救人，只须一个人来到大理，也不能让他们走 | perplexity 19.756388  他心中一凛，暗想这人武功虽高，却是一招之间，却是不能。这时他武功虽高，但武功高强，却是一招之间，竟然 |
| 110 | perplexity 19.503972  ，不敢再行凑直。他们是个女子，不敢有丝毫不厌功劳。他们是个女子，不敢有丝毫疑心，但见他脸上一红，心中 | perplexity 19.503972  ，在他手腕上一拍，一股劲风扑到。段誉心想这小子武功高强，不能再理会这些人，这是我的大大的手法。这一下 | perplexity 19.647875  他心中一动，道我不知道，我不知道，你这小子是我的好朋友，我也不能跟你说。我不知道，你不知道，我不知道 | perplexity 19.647875  洪七公大惊，叫道你这小子是我的功夫，你不能让他们逃走，那就是了。郭靖道我不是你的手，我也不会。黄蓉道 |
| 120 | perplexity 19.403819  ，这一下再也不能动弹。这一下变起仓卒，一颗心也不禁大为惊讶，一时说不出话来。那人道你们这么一个人，你 | perplexity 19.403819  ，在他手腕上一点，手指一搭，已抓住了他左腕，右手一扬，两人都已将他身子带上，向段誉胸口撞去。段誉心中 | perplexity 19.528124  他心中一凛，暗想这女子是谁？那女子道我们是大理国的皇帝，不知是甚么意思。段誉道我不知道，我不知道，你 | perplexity 19.528124  他心中暗惊，暗想这人武功虽高，却也不能再使这样的功夫。这一招，一招之下，竟是一招亢龙有悔，右手一招， |