**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**

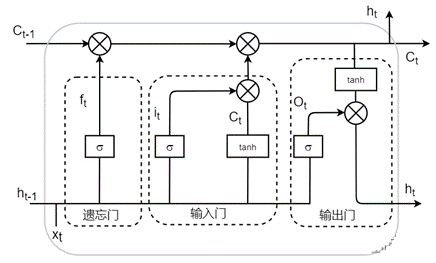


图 1 LSTM

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)。

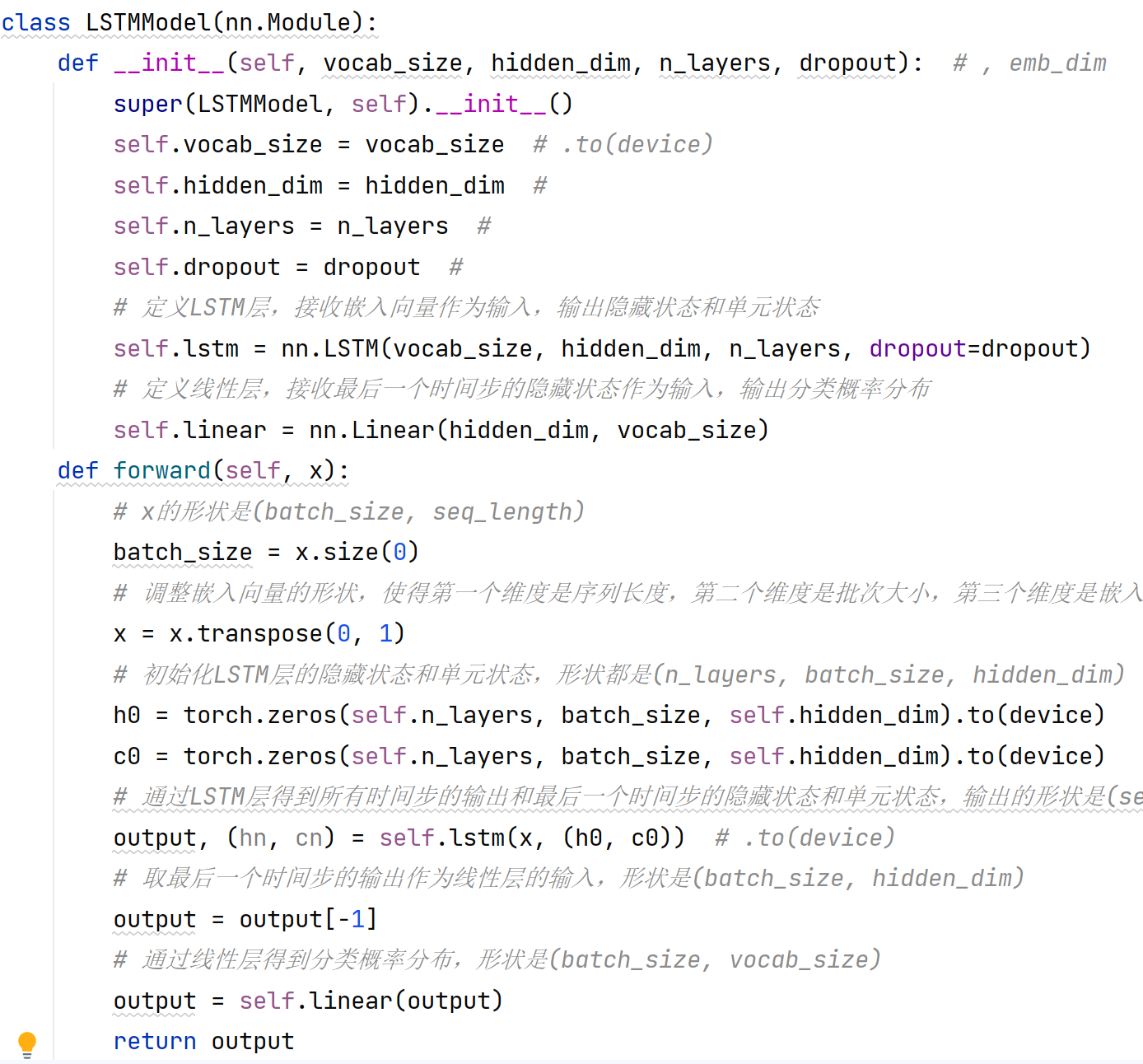
本次代码使用了Gensim，一个简单搞笑的自然语言处理库中的Word2Vec算法，将经过预处理后的文字进行向量编码作为LSTM算法的输入，通过对LSTM算法输出的向量与word2vec模型中的向量进行比较，选取相似度较高的字作为输出进行解码。

**Methodology**

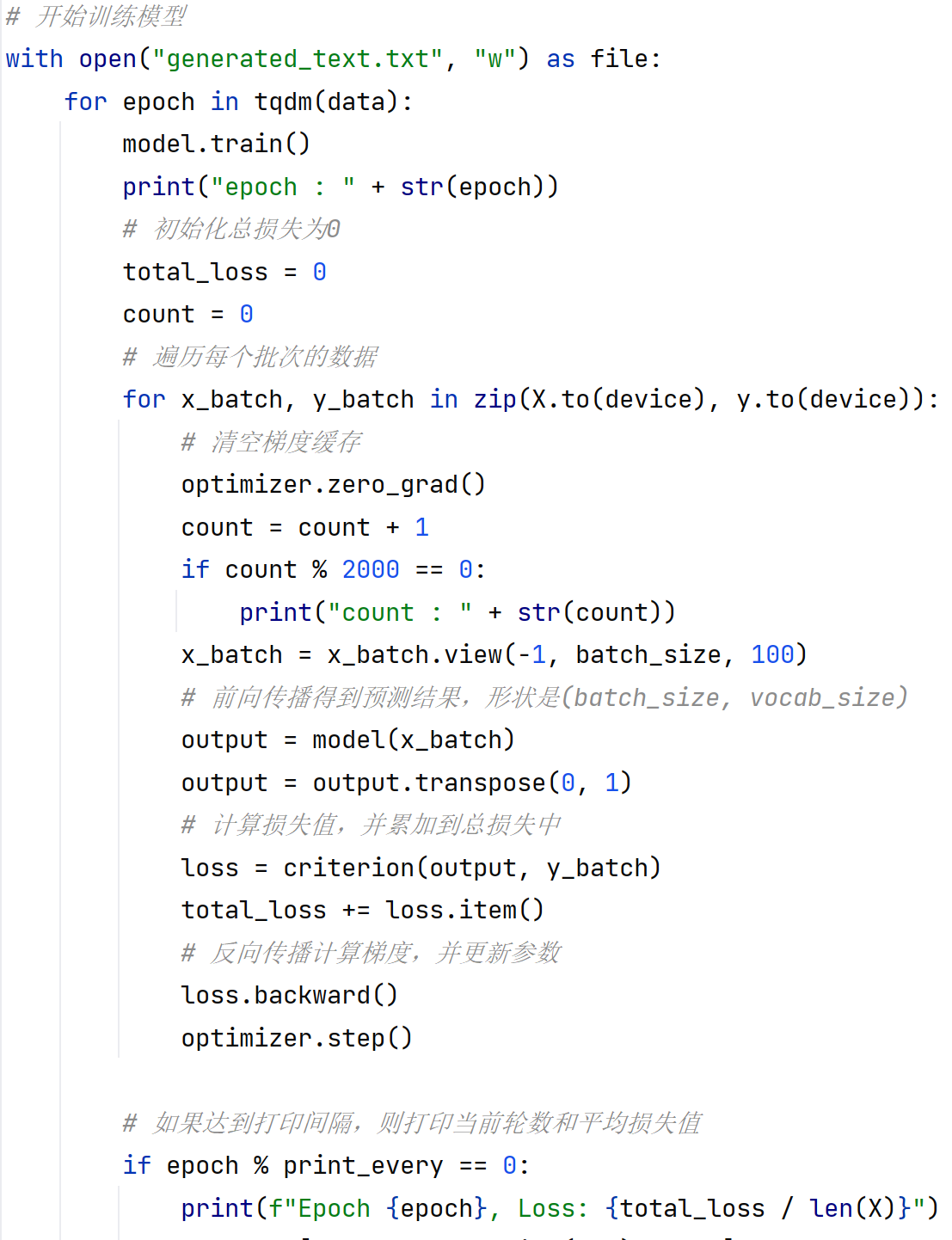
首先对文本进行预处理然后通过gensim的word2vec将文本中文字转换为256维的向量进行表示。



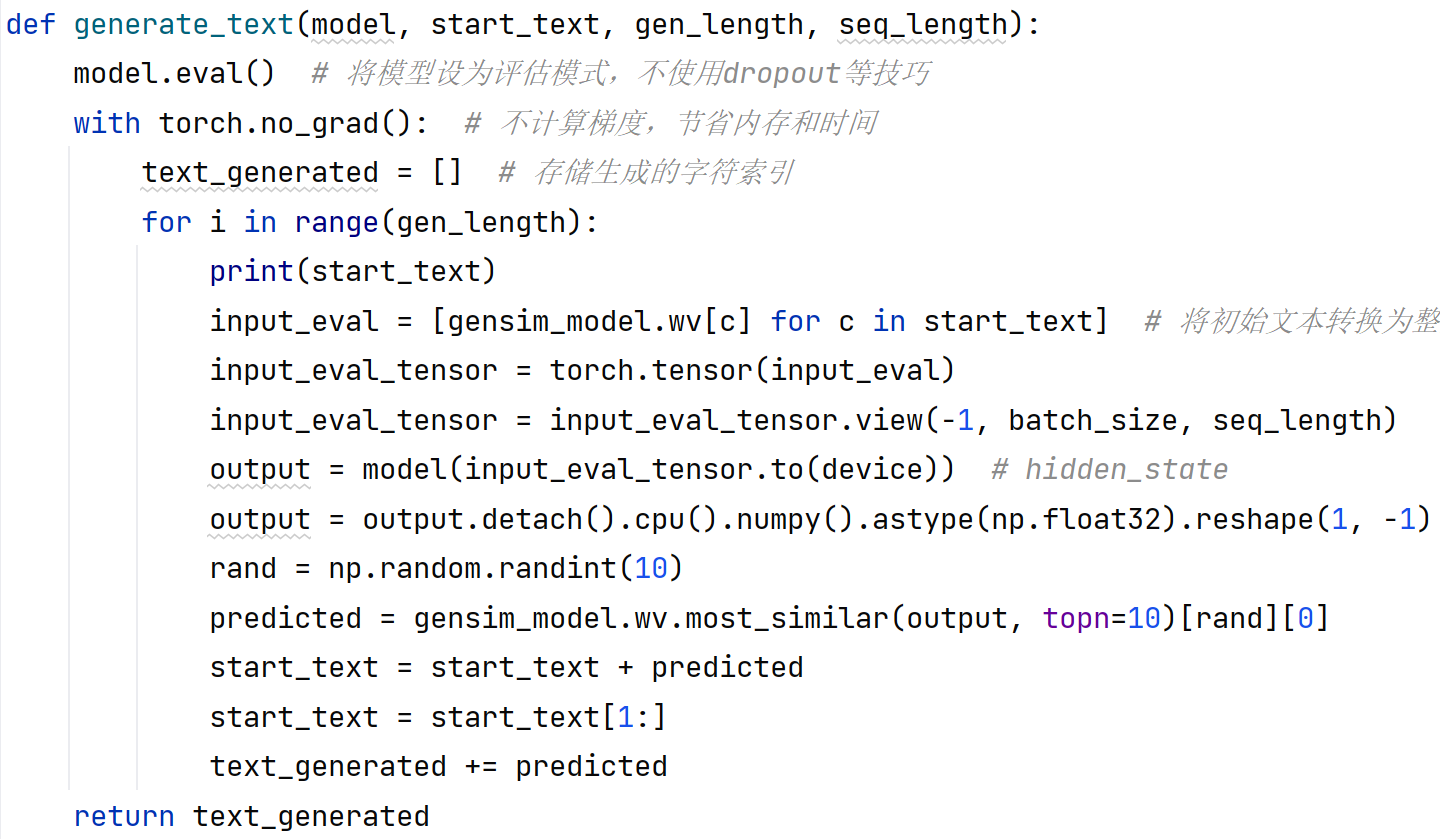
然后对LSTM网络进行定义



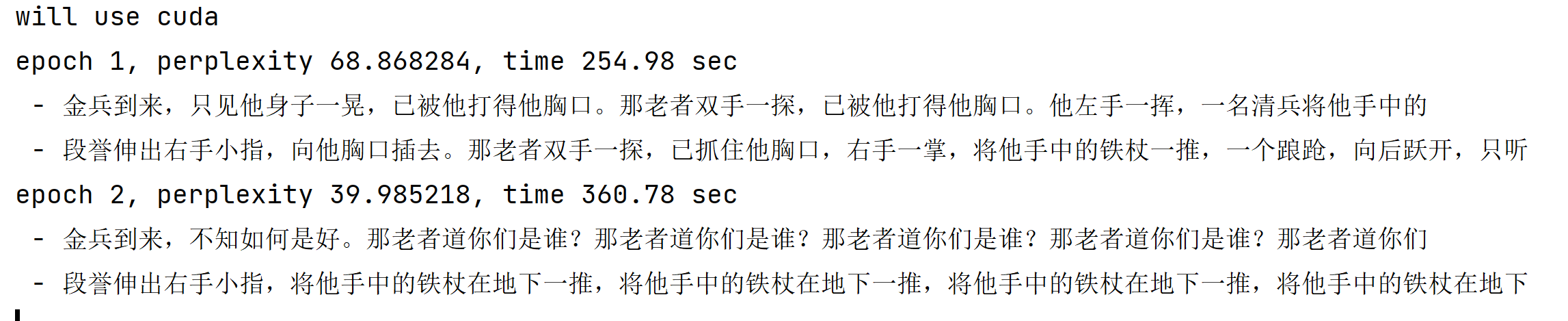
将数据库文本转为向量后，按照每20个一组的顺序依次输入输入网络，对下一时刻的文字向量进行预测，并与真实文字的向量进行比较。损失函数使用mse函数。对模型进行训练：



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：



epoch 1, perplexity 69.225284, time 362.43 sec

- 他立刻闭眼，心中怦怦乱跳，过了片刻，才慢慢睁眼，只见所处之地是座圆形石室，光亮从左边透来，但朦朦胧胧地不似天光。 这一下，那是个女子，不知是甚么东西？那老者道你们不知道，我们不知道。那老者道你们不知道，我们不会说，

- 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。 那老者叫道好，你们这个小子，你们不知道？那老者道你们这个小子，你们不知道？那老者道你们这个小子，你们

epoch 2, perplexity 40.158431, time 362.66 sec

- 他立刻闭眼，心中怦怦乱跳，过了片刻，才慢慢睁眼，只见所处之地是座圆形石室，光亮从左边透来，但朦朦胧胧地不似天光。 他心中一凛，只见他身形一晃，已将他手中长剑夺过，将他手掌翻了出来，将他手中长剑夺过，说道你们这么一拳

- 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。 这一下变得甚快，不敢贸然上前，只听得嗤的一声响，一个女子的身躯向前飞出，一个踉跄，向前跨了一步。那老