问题描述

基于 Seq2seq 模型来实现文本生成的模型，输入可以为一段已知的金庸小说段落，来生成新的段落并做分析。

实验原理

Seq2Seq

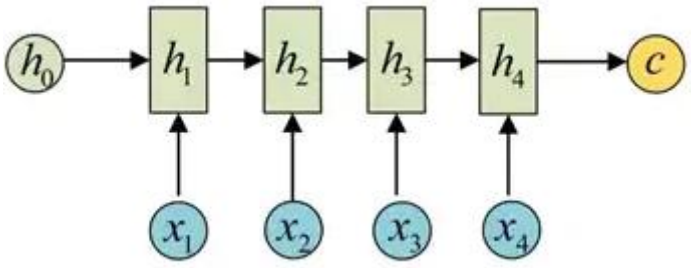
Seq2Seq 模型是输出的长度不确定时采用的模型，这种情况一般是在机器翻译的任务中出现，将一句中文翻译成英文，那么这句英文的长度有可能会比中文短，也有可能会比中文长，所以输出的长度就不确定了。如下图所，输入的中文长度为 4，输出的英文长度为 2。

在网络结构中，输入一个中文序列，然后输出它对应的中文翻译，输出的部分的结果预测后面，根据上面的例子，也就是先输出“machine”，将"machine"作为下一次的输入，接着输出"learning",这样就能输出任意长的序列。

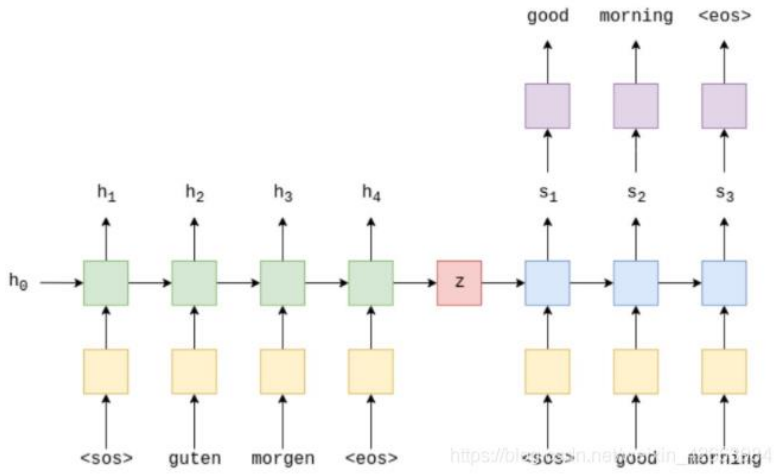
机器翻译、人机对话、聊天机器人等等，这些都是应用在当今社会都或多或少的运用到了我们这里所说的 Seq2Seq。

Seq2Seq 结构

seq2seq 属于 encoder-decoder 结构的一种，这里看看常见的 encoder-decoder 结构，基本思想就是利用两个 RNN，一个 RNN 作为 encoder，另一个 RNN 作为 decoder。encoder 负责将输入序列压缩成指定长度的向量，这个向量就可以看成是这个序列的语义，这个过程称为编码，如下图，获取语义向量最简单的方式就是直接将最后一个输入的隐状态作为语义向量 C。也可以对最后一个隐含状态做一个变换得到语义向量，还可以将输入序列的所有隐含状态做一个变换得到语义变量。



首先将源语句输入至 encoder 编码为一个向量，我们称为上下文向量，它可以视为整个输入句子的抽象表示。然后，该向量由第二个 LSTM 解码，该 LSTM 通过一次生成一个单词来学习输出目标语句。下面给出文本翻译的例子。



源语句被输入至 embedding 层（黄色），然后被输入编码器（绿色），我们还分别将序列的开始（）和序列的结束（）标记附加到句子的开始和结尾，sos 为 start of sentence，eos 为 end of sentence。在每一个时间步，我们输入给 encoder 当前的单词以及上一个时间步的隐藏状态 h_t-1，encoder 吐出新的 h_t，这个 tensor 可以视为目前为止的句子的抽象表示。这个 RNN 可以表示为一个方程： ht=EncoderRNN(emb(xt),ht-1) 这里的 RNN 可以是 LSTM 或 GRU 或任何 RNN 的变体。在最后一个时间步，我们将 h_T 赋给 z，作为 decoder 的输入。

在每个时间步，解码器 RNN（蓝色）的输入是当前单词的嵌入，以及上一个时间步的隐藏状态，其中初始解码器隐藏状态就是上下文向量，即初始解码器隐藏状态是最终编码器隐藏状态。因此，方程为 s_t=DecoderRNN(emb(y_t,x_t-1),然后在每一个时间步，我们将 s_t 输入给线形层（紫色），得到 y_t_hat,即 y_t_hat=f(s_t) ,而后用 y_hat 与 y 进行交叉熵计算，得到损失，并优化参数。

Encoder(编码器)

在前向计算中，我们传入源语句，并使用嵌入层将其转换为密集向量，然后应用 dropout。然后将这些嵌入传递到 RNN。 当我们将整个序列传递给 RNN 时，它将为我们自动对整个序列进行隐藏状态的递归计算。请注意，我们没有将初始的隐藏状态或单元状态传递给 RNN。

Decoder(解码器)

解码器执行解码的单个步骤，即，每个时间步输出单个 token。 第一层将从上一时间步中接收到一个隐藏的单元格状态，并将其与当前嵌入的 token 一起通过 LSTM 馈送，以产生一个新的隐藏的单元格状态。 后续层将使用下一层的隐藏状态，以及其图层中先前的隐藏状态和单元格状态。

Decoder 的参数与 encoder 类似，但要注意它们的 hid_dim 要相同，否则矩阵无法运算。

模型应用领域

首先作为为机器翻译问题为出发点提出来的 seq2seq 模型，机器翻译的准确率因为该模型的提出而有了较大的提升。

作为 seq2seq 模型研发团队，Google Brain 团队在 2014 年的文章的应用案例中对 LSTm 的隐藏结点做了主成分分析，如下图所示，从图中可以看出，模型中的语境向量很明显包涵了输入序列的语言意义，后沟将不同次序所产的的不同意思的语句划分开，这对于提升机器翻译的准确率很有帮助。其次 seq2seq 模型因为突破了传统的固定大小输入问题框架，因而除了翻译场景，还被用于智能对话与问答的实现以及微博的自动回复，2015 年华为团队，通过 seq2seq 为基础设计的模型实现了计算机对微博的自动回复，并通过模型间的对比得到了一系列有意思的结果。如下图，post 为微博主发的文，其余四列为不同模型对该条微博做出的回复。

ID	Post	NRM-glo	NRM-loc	NRM-hyb	Rtr.-based
1	每年元旦都来高烧。 High fever attacks me every New Year's Day	多休息，增强免疫力！ Rest more to boost your immune system.	每年元旦都去医院看病 Go to see a doctor every New Year's Day	祝早日康复，身体健康！ Get well soon and stay healthy!	高烧… High fever…
2	又老了一岁，感谢我们组，我很高兴 I gain one more year. Grateful to my group, so happy	生日快乐！永远支持你！ Happy birthday! Will stand by you forever!	生日快乐！永远年轻漂亮！ Happy birthday! Wish you always stay young and pretty!	老了老了，岁月不饶人啊！ Getting old now. Time has no mercy.	我只是老了一岁 I just gain only one more year.
3	吃素第一天，坚持住，崔朵拉。 First day of being a vegetarian. Hold on, **CuiDuoLa**	必须来支持！加油！ I come to offer my support. Keep it up!	坚持住，坚持下去。 Hold on, keep it up.	每天都是素食主义者 I have been being a vegetarian everyday	坚持住！ Hold on!
4	这种天气真应该带个帅锅出去踏青赏花 We should go out with some cute guys to enjoy a great outing in such a nice weather.	这是哪儿啊？真美！ Where is it? It is so beautiful!	阳光明媚，心情舒畅。 Such a nice sunny day! I am in a great mood.	这种天气真不是盖的。 It is indeed a very nice weather.	文山啊 出去踏青寻找灵感 哈哈 **WenShan**, let's go out to get some inspiration. Ha! Ha!

实验结果

```
####################
Temperature: 0.2
####################
    杨过见她脸色苍白，眼见她泪盈眶水，滚夜滚开，滚来滚去，滚夜抖动，滚开绸带，绸带末端的拆解开来。绸带拆解，带著金铃击倒，甩开解开。李莫愁拂尘挥动，拂尘卷起土来，拂尘往她脸上一卷，宛如一块大树皮结唤，脸上似是一丝笑道：「你若是我爹爹知道？」

    杨过见她这般温柔诚恳，心中一动，道：「你说我是媳妇儿，我教你的心意，我要你做我的。」

    杨过道：「我不知道，你不是我的。」那少女道：「我跟你说，你是谁？」陆无双道：「你要我杀你，我也不会跟你说。」杨过道：「我不知道，你不是我。」

####################
Temperature: 0.5
```

```
####################
```
杨过见到这少年，心中大喜，道：「你不用跟我说话，我跟你说，我在这里陪著你。」杨过道：「我不知道。」洪凌波道：「你瞧我，我就在这儿干麼。」

那少年不见怪异，杨过正自沉吟，心想：「这姑娘是谁的？」说著站在天井之中，最後仰天便往地底。两个人说不过半枚，直天往北星扑上去。」

杨过道：「你这般无礼，只是我义父师父，你就不能再好。」小龙女道：「我瞧著你的，是甚麼？」杨过道：「你就是我，我也不会。」顿了一顿，说道：「我不知道，你怕我们师父麼？」郭靖道：「为甚麼？」杨过道：「我在这里内功力深。」黄蓉道：「我知道？」黄蓉道：「他定然不知在这儿，重阳宫该当遭劫，若不是她不幸，误会，你若不会打过我，还会下面子山上，难道你就是一样。」

```
####################
Temperature: 1.0
####################
```
那个嘴角带了来，轻著这一跃进了。杨过给她治病，刻身朝负著胆子；照料得到原期竟然不同小龙女。他想到一时所莫错杀，没能尽化解救之下，说道：「小王重阳临危，难道是万祸仙庄子！」郭靖听了郭靖此事说，黄蓉当杀他妻子之口，说道：「难道咱们上当真教。她傻蛋，你说这娃儿罢？」黄蓉拂尘一起，轻身黄蓉再度，见杨过站在她身後。郭靖呆出房门外，慢说：「九阴真经功夫。」

杨过瞧她一时，难以与不成的小龙女并肩而来，柔声道：「唉，果然极是温柔，从姑娘生想取了这一天真下对敌人的手帕有来造诣。此时若是别多，温言反而便得到万事。表姊、傻是傻子聪明。郭靖本来就撞连几线、程陆逊去，陆家脱身双姊姊妹，时见疯疯疯癫，怕己傻疯却乘、狂以之势骇死。她心里傻笑，疯疯癫癫癫癫，或忘迎像像前洪老矣、双仞，靖、黄蓉上两、一灯、大师、一灯、周伯通等四周大遭大的关式都是攻的藏在周身前阁，身的一脚实已易躲，不敢再瞧到他们去见啦。

小龙女见他过身摇幌，扑击过地，不似招术。玉女剑法自奇，那时石顺畅饮，凑眼巧落，突觉如此落人，手足向武修文身旁急道：「过儿，那位了姓龙的书生，你已经过他的那万不及，武功法测儿。」李莫愁对付诸流人，其後必定夹攻，自他二人也难练了，不过他们急忙。黄药师道：「我……你一听说的。」店小二女说过饭菜，又问受甚麼连公孙止了？一人叫道：「朱伯伯，我师父有两位和甚师兄为难相似，拉住他头发便深入了寸风。」杨过笑道：「郭伯伯，我武家哥哥武功不弱，此时还得虽有没能让他们胡子人力，我们却有了何师我忍气？因之师父丝毫不剩下气，又何必为争？」脸色苍苍老迈，那老丐如何黯然不然足力老衰，又何以之又变招法精降专研心的武功？

```
Process finished with exit code 0
```