《深度学习基础》课程大作业

|  |  |  |
| --- | --- | --- |
| 具 体  题 目 | 基于seq2seq的对联生成 | |
| 学 号 | 姓名 | 成绩 |
| **55190828** | **张天浩** |  |
| **55190831** | **唐 帅** |  |
|  |  |  |

# 一、 内容介绍

## 1.1 简述介绍题目的动机、目的和主要内容

**题目动机：希望通过大量对联,完成一个可以通过上联，自动完成下联的模型；**

**实验目的：学习seq2seq问题的解决方法，学习编码器-解码器框架，对模型进行调试和测试，并对模型进行适度更改。**

**主要内容：包括编码器——使用LSTM实现，解码器使用LSTM + attention机制，以及特殊的预测网络。**

# 二、解决方案及细节

##### 2.1 对联

对联的上下联是一个典型的序列到序列建模的场景，编码器-解码器（ Encoder - Decoder）框架是解决seq2seq问题的经典方法，它能够将一个任意长度的源序列转换成另一个任意长度的目标序列：编码阶段将整个源序列编码成一个向量，解码阶段通过最大化预测序列概率，从中解码出整个目标序列。编码和解码的过程通常都使用RNN实现。

##### 2.2 句子标准化

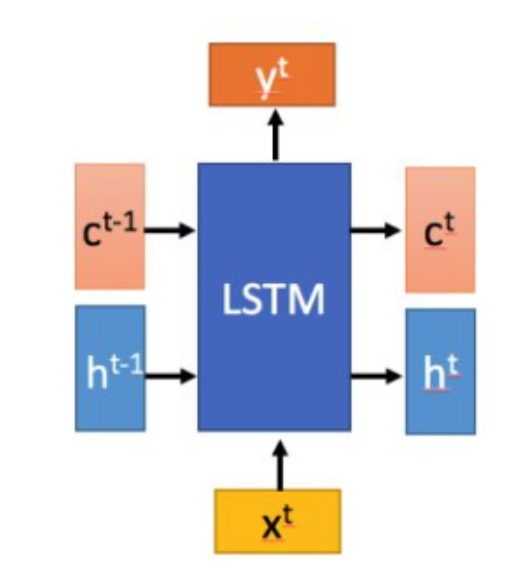
将对联进行编码（如one-hot编码）并通过padding、堆叠后得到具有统一格式，完成句子标准化

##### 2.3 Encoder

###### 2.3.1 embedding层

通过输入的句子编码得到一个句子长度\*词向量维度的词向量，以作为下一层网络的输入。

###### 2.3.2 LSTM层



通过h,c两个状态值保存每一步的结果，并最终由h,c的过程量得到output和最后的h\_n,c\_n。

##### 2.4.Decoder

###### 2.4.1 AttentionLayer

AttentionLayer内输入编码器的输出序列和解码器的输出序列，并将其作为 softmax层的输入，最后得到注意力权重。

###### 2.4.2 Seq2SeqDecoderCell

作用是作为Decoder的处理单元，需要在这里构建具有attention的LSTM，里面需要LSTMcell作为基础，完成n层LSTM网络的设置，最后得到网络的输出和状态值（h,c)。

###### 2.4.3 Seq2SeqDecoder

使用nn.RNN，将输入的Cell封装为一个循环神经网络。它能够重复执行 cell.forward() 直到遍历完input中的所有Tensor。Seq2SeqDecoder的主要功能为根据编码层的输入完成对于下联编码的预测。

##### 2.5 参数选择

优化器：Adam

loss:交叉熵损失函数

学习率：0.001

编码器LSTMdropout：0.2

最大轮数：20——在不同轮数结果不一致

##### 2.6 模型预测——Seq2SeqAttnModel

继承主函数，作为预测函数，输出预测的下联序列，最后使用字典将序列转换回文字即可。

# 三、改进策略

##### 3.1 你提出了什么样的改进策略，效果如何

###### 3.1.1 修改参数

参数设置如下图所示：



其中max\_grad\_norm用于控制梯度膨胀；learning\_rate表示学习率；max\_epoch表示训练轮数；log\_freq控制日志打印输出；batch\_size表示每批数据的规模大小。

###### 3.1.2 训练结果

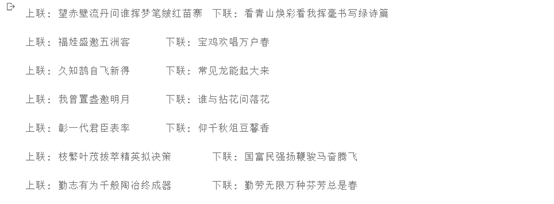
最终loss值为1.1156，Perplexity值为29.6325，可以看出得到了较好的效果。其中Perplexity表示困惑度，用于评价语言模型，其值越低表示模型效果越好。

# 四、 实验结果分析

##### 4.1 哪些因素导致了当前的实验结果？

###### 4.1.1 实验结果

  部分实验结果如图。对联的基本要求为字数相等、断句一致、要平仄相合、音调和谐、词性相对、位置相同、内容相关、上下衔接等。分析实验结果，输出的上下联基本满足这些要求。如风云对岁月，漫天瑞雪对遍地春风，千山秀对万里香等，且语义上上下联都有衔接。因此可以判断本次构建的编码器-解码器模型在对对联的应用上取得了良好的效果。



###### 4.1.2 因素

  ①编码器-解码器(Encoder-Decoder)框架的应用。能否选取正确的、合适的框架很大程度上决定了一个问题能否被正确地、快速地、合理地解决。一个良好的应用框架对解决一个问题有着事半功倍的效果。本项目中对对联是一个典型的序列到序列(sequence2sequence, seq2seq)问题，而编码器-解码器框架是处理seq2seq问题的经典办法。其中编码器将输入序列编码为中间信息，解码器将中间信息解码为目标输出。这类框架在其它方面也有广泛的应用，如机器翻译、文本摘要、图像描述生成、图像生成等。

②采用LSTM网络分别实现编码器和解码器。各种网络都有自己的优缺点，因此不同的网络分别有各自的应用领域。例如CNN网络，它的权值共享网络结构使之更类似于生物神经网络，降低了网络模型的复杂度，减少了权值的数量。该优点在网络的输入是多维图像时表现的更为明显，使图像可以直接作为网络的输入，避免了传统识别算法中复杂的特征提取和数据重建过程。因此CNN网络被广泛应用在图像识别领域：RNN网络中它每次都会将前一次的输出结果，带到下一次的隐藏层中，一起训练即能够认识到之前的输入。因此RNN网络被广泛用于处理序列问题，如语音识别、文本生成等。针对如何认识之前的输入RNN衍生出了另外几种不同的网络，如我们提到的LSTM。LSTM能够保存序列中重要的信息，忽略其它信息，由此解决了RNN网络的短期记忆问题。保存重要信息的这一特点刚好非常契合文本序列问题。

③注意力机制的应用。注意力机制可以简单的理解为一种分配权重、分配资源的方式。以本项目中的编码器-解码器为例，文本序列输入编码器，编码器产生中间输出，该中间输出与解码器其它输入一起输入到解码器中，解码器得到输入后产生目标输出，目标输出下一次会作为解码器自身输入的一部分，以此为循环，这就是编码器和解码器大致的工作步骤。其中编码器的输出不会因为解码器之前的输出内容而做相应改变，没有考虑到两者之间存在的联系。加入注意力机制后，每次依据解码器输出得到的权重会作用于编码器的输出，这样就构建起了编码器输出与解码器输出两者内容的联系，更利于网络的训练。

##### 4.2 还可以从哪些方面进行改进从而获得更理想的实验结果

首先是参数的设置。这里的参数指学习率、批次大小、训练次数以及优化器、激活函数的选择等这一系列配置信息。因大部分神经网络都需要大量重复训练的特点，除了正确框架的搭建，合适网络的采用，良好参数的设置对一个神经网络能否有不错的结果也起着至关重要的作用。拿其中的学习率这一个参数来说，学习率控制我们要多大程度调整网络的权重，以符合梯度损失。学习率设置过大，网络无法找到最优值，使其无法收敛；设置过小网络收敛缓慢，大幅度增加了训练网络的时间。好的参数设置能让神经网络达到极佳的效果，劣质的参数设置不仅可能使网络表现差，甚至可能在网络的训练过程中根本得不到良好的训练结果，使一切的工作变得毫无意义。但目前为止，参数的设置还主要以经验为主，没有什么系统的规则或方法。

其次是实现编码器-解码器框架的网络的选择，这里被选中的网络是LSTM。并不是所有的编码器-解码器框架都要用LSTM网络实现，针对于不同的应用场合可以选择不同的网络。上文提到LSTM网络能够保存序列中重要的信息这一特点非常契合文本序列问题，但并不是只有LSTM才能具有这样的特点。其它的网络结构经过一定变形或者LSTM自身再经过一定的改变而形成的新的网络结构可能会有更好的效果。

# 五、 总结

## 5.1 介绍每人承担的具体工作

二人均完成了对实验的学习和调试。

张天浩完成了实验报告第一节内容介绍，第二节解决方案及细节和第五节总结的部分。

唐帅完成了实验报告第三节改进策略及第四节实验结果分析的部分。

## 5.2 请总结本次大作业的收获

在本次大作业中，我们完整地学习了seq2seq问题的解决方法，并且在学习的过程中了解了LSTM,attention模块的原理与结构。这个大作业虽然并非图像识别方向的实验，但是在实验中，我发现二者之间有异曲同工之妙，二者均是转换为tensor类型作为输入，输入后通过模型进行训练，并计算loss进行收敛，最后使用计算预估值的函数进行预测。在课程中，我们学到如何使用最基础的MLP，CNN完成一次训练，也学到了使用更复杂的Resnet网络利用残差网络减少信息损失。这次大作业和课程给我最大的收获是教会了我如何通过已经学过的大致框架去学习一种新的模型。这个课程虽然只完成了少量的项目的学习，但确实起到了师傅领进门的效果。

## 5.3 请给出对《深度学习基础》课程的建议[footnoteRef:2]

希望能够增加课程内对项目代码进行详细阐述的实践课，可以减少许多实验中会出现的疑惑。