assignment3 我的答案

xzx

2016 年 12 月 7 日

1 Q1: Image Captioning with Vanilla RNNs

1.1 Recurrent Neural Networks

首先是单个神经元的前向与后向传播。一个神经元的示意图如图 1所示.

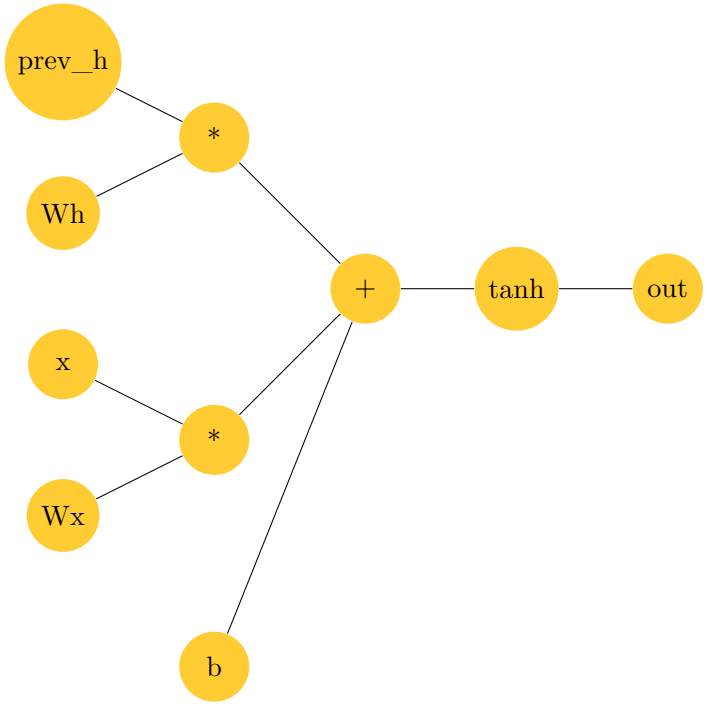

图 1: rnn_step_forward

反向传播也很容易根据这个图示推导出来.

整个 rnn 的前向与后向传播再课程里面已经把图给话出来了。这里面 Wh 还有 Wx 被重复使用了很多遍。对于一个变量后分成多条路径的，把这几条路径上面的微分相加。

1.2 RNN for image captioning

rnn 的输入有点多，看上去有些混乱，整理一下输入放在图3 中可以看的清楚一些。以图 2为例作为输入.

图 2: sample_input

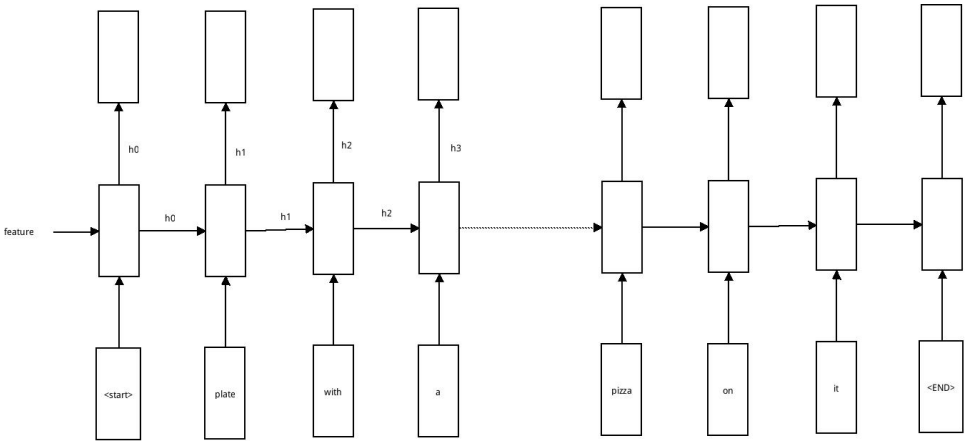

图 3: RNN for image captioning

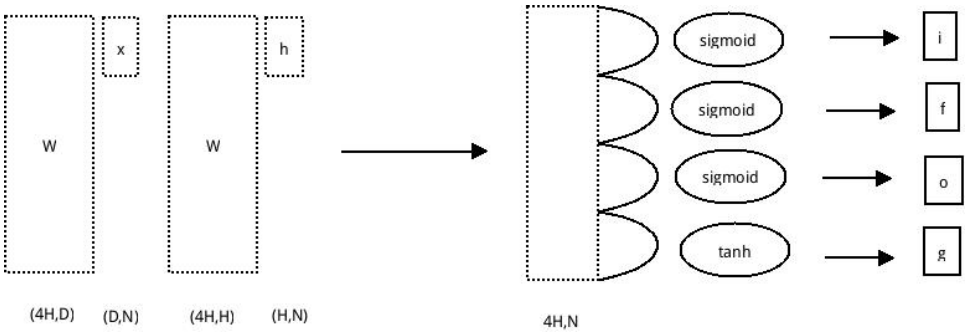

图 4: LSTM cell

每个神经元的输出会同时传给下个神经元并且向上传到上一层网络。这个 ipynb 的最开始就说已经帮我们把图像的特征提取出来了，因此 feature 当做最初的 h0 传入网络就好了。每个单词对应的向量作为 x 从下面输入到网络。

2 Q2: Image Captioning with LSTMs

人类在进行学习的时候，实际上会忘掉一些东西。比如一些东西很久没有通过了就会忘掉。LSTM 就是 rnn 在这个基础上面的变化。

课程的 PPT 里有几个从论文里面摘下来的图，我觉得这几个图不是很好理解，我在这几张图的基础上做一些改变，结合作业中的代码希望能够更清楚的解释 lstm。下面都默认忽略掉 bias，这样画图能简单些。

如图4，输入的 x 跟 hidden layer 跟原来差不多，输出从原来的 H 变成了现在的 4H。要把这个 4H 平均分成四份，每一份代表了一个门。i 表示输入;f 表示 forget，就是前面说的要忘记一些东西；o 表示输出；g 表示什么视频中说他也不知道，反正就是一个门。然后下一层的 h 要有这些一起来决定。公式 ppt 里都有。

这样，标准的 rnn 的图就改成了 ppt 第 68 页的图，原来的 hidden layer 只有一个绿色的框，现在在绿框下面增加了一个黄色的框，就是公式里面的 c，表示 cell。

把模型实现了以后可以看到它会学习看图说话。有些图片说出来的话挺搞笑的。

3 Q3: Image Gradients: Saliency maps and Fooling Images

这一节主要的想法是利用反向传播修改原图片，假如微小的噪声后使原图片被误分类。ImageGradients.ipynb 中数据加载要 2.8GB 的内存，所以建议如果电脑的内存只有 4G 的话，尽量先去把代码写好，然后再加载数据。尽量加载一次数据多做一些事情。反正我每次都有大概十分钟浪费在加载数据上面。

3.1 Saliency Maps

题目中说要先去读一读参考文献的第二章, 这篇论文百度就可以找到。(这一块作业好像写的有点问题，容我再看看。。)

3.2 Fooling Images

这一小段的目的就是要把原来的图片做细小的更改，使得 cnn 会误分类。

首先把图片正常的前向传播，会得到一个当前图片的分类。因为用的都是筛选过的 validation 的数据，所以第一次得到的分类肯定是正确的。假设正确的分类是 a，目标是把产生一个相似的图片，使得误分类为 b。那么第一次前向传播得到的结果肯定是 a，然后将 a 与 b 之间求损失函数的值，再反向传播回去。模型中都会输出一个 dx，把原来的图片减去 dx 就好。有可能一部分像素被减成负的了，在输出图片的时候取绝对值输出。