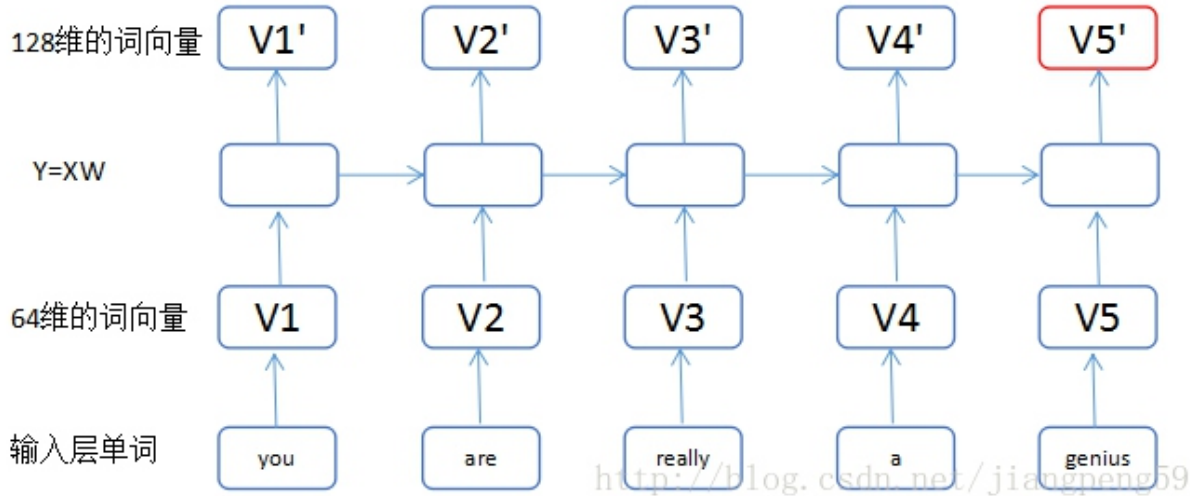
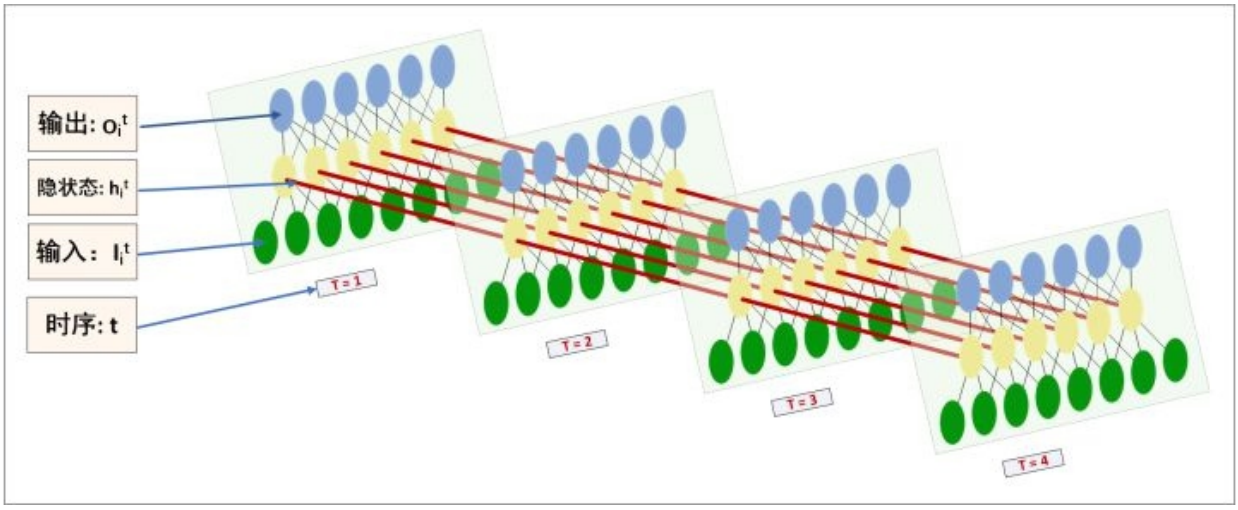
RNN的输入和输出代表什么意思

对于堆叠多层lstm，lstm的超参和输出的数据的多个维度代表什么意思。

如下是堆叠多层lstm在时间这个维度上的一个截面



如下的多层RNN，总共有三层（输入，隐藏层，输出），如下的时间窗口为4，这4个网络的参数是共享的。单独来看每个时刻的切面，就是一个MLP的增强版：原始的MLP的基础上，上一个时刻的hidden cell的值，需要传给本次timestep对应MLP的hidden cell作为输入。

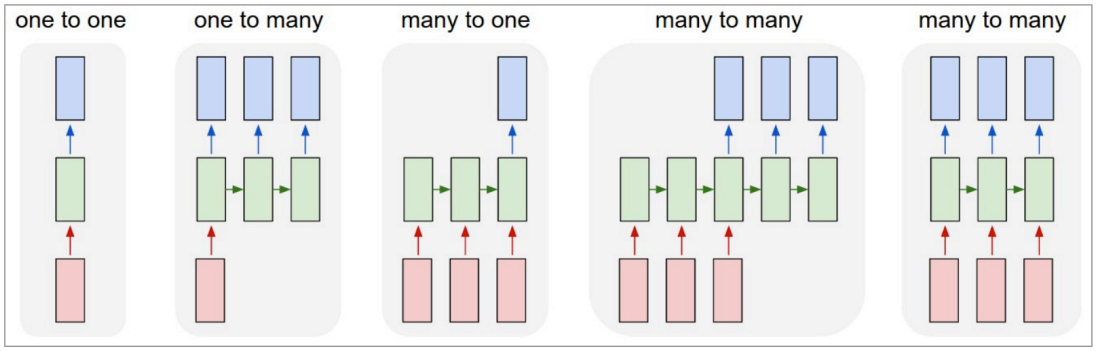


上面讲了rnn的运行机制，至于更复杂的rnn比如LSTM、RGU，那就是在MLP的每个hidden cell（一个黄色circle）与下一个time_step 的hidden cell的传值机制的more sophisticated tactics，based on this text 现在应该好理解了。

RNN的常用结果举例

- 1对1：时间步长为1，退化为MLP, 比如图像分类。
- 1对多：图片取标题。
- 多对1：情感分析/时序分类/时序异常检测。
- 多对多1：机器翻译/多步骤时序预测。
- 多对多2：视频分类，实时的对每一帧画面标记。

注意rnn模型对输入序列的长度没有限制，只要是单个时间长度的整数倍就行了，因为中间的隐藏层是可以使用任意多次。如下是在时间轴上的切片，每个格子代表的是多个cell。



Each rectangle is a vector and arrows represent functions (e.g. matrix multiply). Input vectors are in red, output vectors are in blue and green vectors hold the RNN's state (more on this soon). From left to right: **(1)** Vanilla mode of processing without RNN, from fixed-sized input to fixed-sized output (e.g. image classification). **(2)** Sequence output (e.g. image captioning takes an image and outputs a sentence of words). **(3)** Sequence input (e.g. sentiment analysis where a given sentence is classified as expressing positive or negative sentiment). **(4)** Sequence input and sequence output (e.g. Machine Translation: an RNN reads a sentence in English and then outputs a sentence in French). **(5)** Synced sequence input and output (e.g. video classification where we wish to label each frame of the video). Notice that in every case are no pre-specified constraints on the lengths sequences because the recurrent transformation (green) is fixed and can be applied as many times as we like.

参考资料

1. https://www.zhihu.com/question/41949741?sort=created
2. https://blog.csdn.net/jiangpeng59/article/details/77646186
3. https://github.com/Vict0rSch/deep_learning