harry is boy and harry is very nice

参数共享的作用就是能在第一个位置识别harry出来，那么同样在第五个位置也能将harry识别出来；因为相同的输入和相同的参数，就会有相同的结果；比如说卷积神经网络，卷积参数共享，换个说法，一个固定的卷积参数必然专门识别某种特征，比如说边框，参数共享就说明该卷积能在图片不同的地方探测到该特征；

所以说参数共享本质上就是说明该模块功能完全相同；在任何一个时间步上，细胞的功能是相同的；

当前时间步的状态 ht 是由当前的输入和t-1时刻积累的信息所决定的，

如果积累的信息忽略不计，那么通过当前的输入，得到的特征应该是一模一样的

Rnn学习不到长时信息的原因：

说明：对于RNN为什么会学不到长距离信息，可以说是RNN会遇到梯度消失的问题，梯度消失的问题，本质上还是使用BP算法导致的，也是激活函数的导数小于1导致的。正如上述所讲的，使用RNN时，在每次输入进行优化计算梯度的时候，会侧重于考虑最近的几次输入，以此让损失减小，而比较远的输入，其实对于计算当前输入的梯度已经不起作用了。RNN可以等价于去调整参数，拟合最近的一些序列，比如说输入的是t时刻，那么可能当前的参数就调整为拟合t到t-p时刻的序列，并且p会比较小，也即RNN的学习不会兼顾到很久以前的输入，就学习不到长距离信息。

如果在反向梯度传播的时候，网络无法对某个输入产生特定的影响，那么就可以理解为网络已经对较远输入产生遗忘，因此RNN只能说是短时记忆。

# **双向RNN**

Bidirectional RNN(双向RNN)假设当前t的输出不仅仅和之前的序列有关，并且 还与之后的序列有关，例如：预测一个语句中缺失的词语那么需要根据上下文进 行预测；Bidirectional RNN是一个相对简单的RNNs，由两个RNNs上下叠加在 一起组成。输出由这两个RNNs的隐藏层的状态决定。

