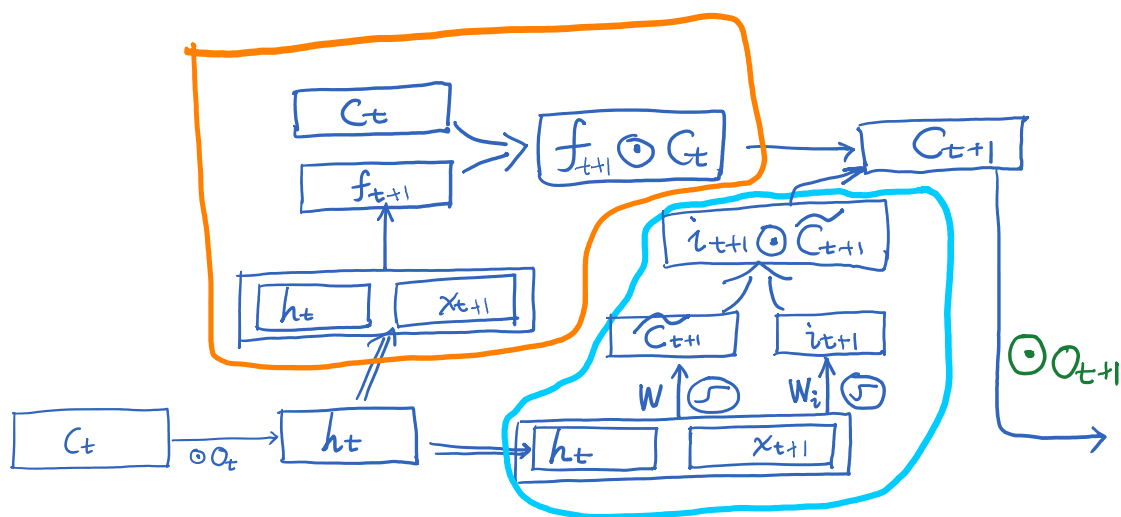


## 图示 GRU 和 LSTM

Sunday, July 23, 2017 1:57 AM

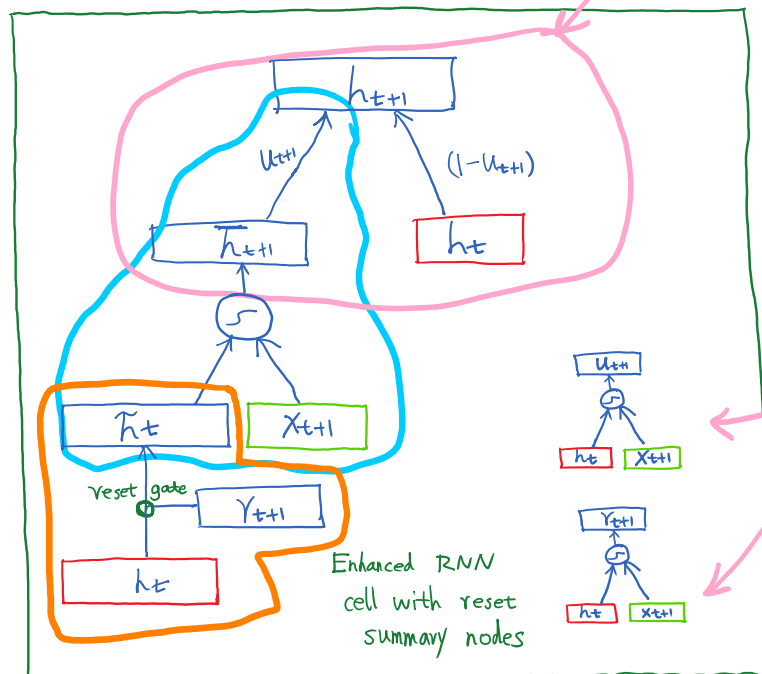


两个模型里面用橙色框起来的部分有些类似，都是  
基于1) 来自上一个时间的输入状态和2) 来自当前时间的输入词语  
一起确定来自上一个时间的输入状态中的哪些维度值得注意  
主要区别是LSTM从上一个时间的h状态决定“选择”（可以类比注意力模型）哪些维度，  
从上一个时间按的c状态“提取”那些维度  
（这里说的“选择”和“提取”通过 $f_t$ 这个gate实现）；  
而GRU不区分c状态和h状态

这个gate可以理解成：

一个维度d要么侧重表达历史记忆（当 $U_d$ 比较小，接近于0的时候）

这个维度d要么侧重表达最新的信息（当 $U_d$ 比较大，接近于1的时候）



两个gate使用  
同样的输入信息计算而来

蓝色部分框起来的部分有些类似

都是（1）基于输入的状态和词语更新状态和（2）产生一个输出用的filter来确定自身状态有哪些值得保留