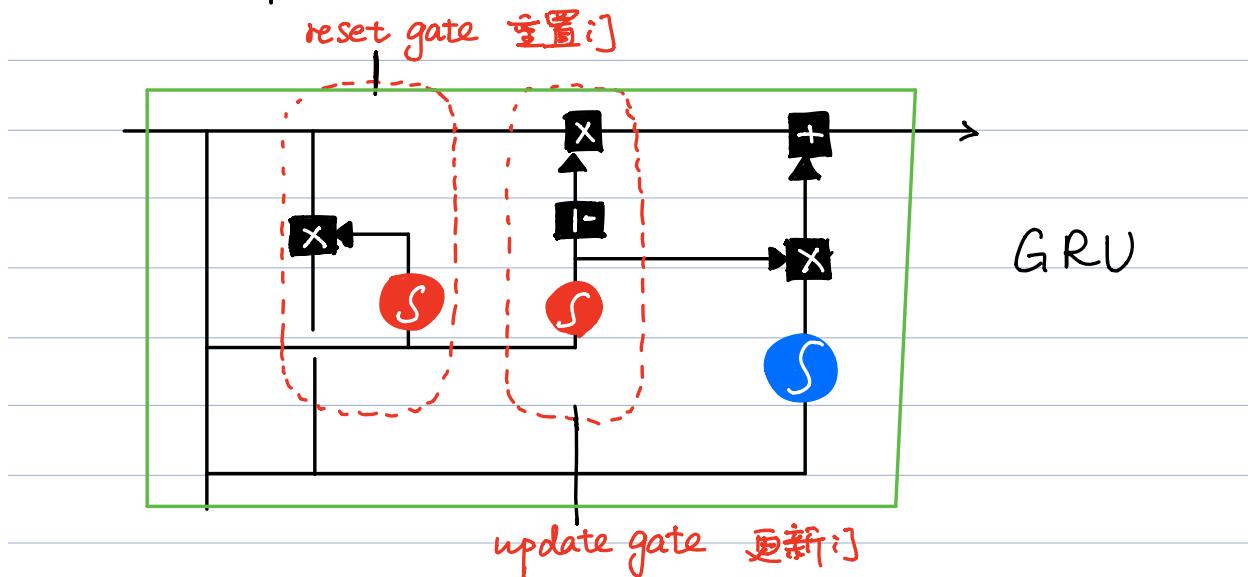


GRU: Gated Recurrent Unit

LSTM 和 GRU 是解决短时记忆问题的解决方案，具有称为“门”的内部机制，可以调节信息流



sigmoid

tanh

pointwise multiplication

pointwise addition

vector concatenation

“门”可以知道序列中哪些重要的信息是需要保留的，而哪些是需要删除的，随后，它可以沿着长链序列传递相关信息以进行预测

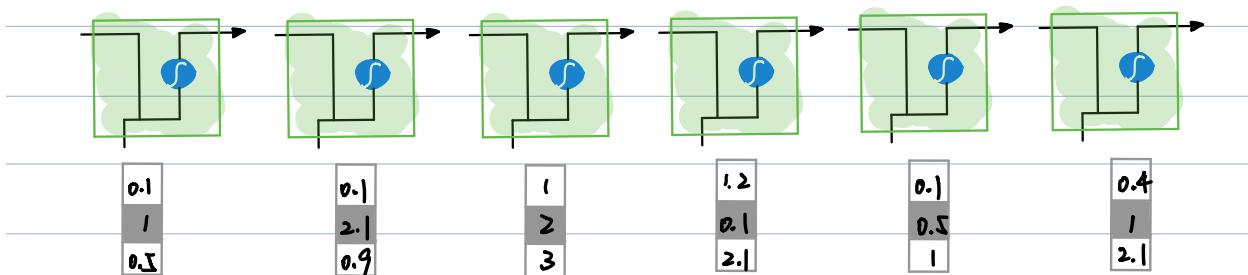
e.g. Amazing! This box of cereal gave me a perfectly balanced breakfast, as all things should be. I only ate half of it but will definitely be buying again!

GRU 可以学习只保留相关信息来进行预测，并忘记不相关的数据

回顾递归神经网络 (RNN) :

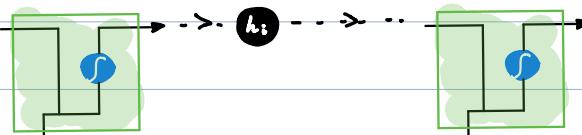
第一个词被转换成了机器可读的向量。

然后 RNN 逐个处理向量序列



逐一处理矢量序列

处理时，RNN 将先前隐藏状态传递给序列的下一步，而隐藏状态充当了神经网络记忆，包含相关网络之前所见过的数据的信息。



⑤ Tanh function ⑥ hidden state (memory)

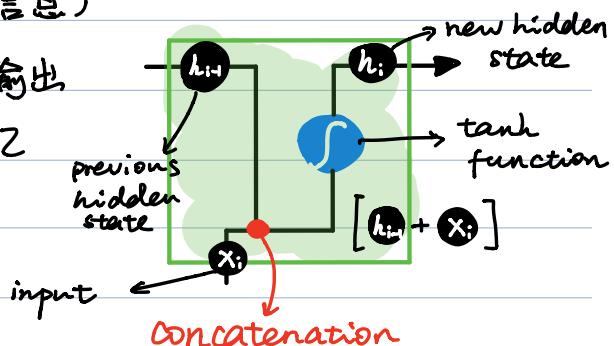
将隐藏状态传递给下一个时间步

RNN 的一个细胞：首先将输入和先前隐藏状态组合成向量。

(包含当前输入和先前输入的信息)

向量经过激活函数 tanh 之后，输出

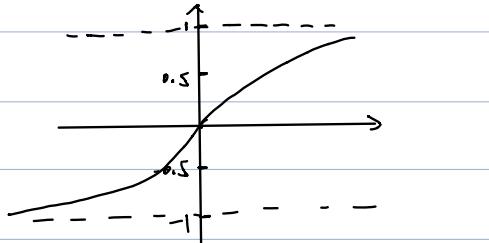
的是新的隐藏状态或网络记忆



激活函数 Tanh

用于帮助调节流经网络的值。

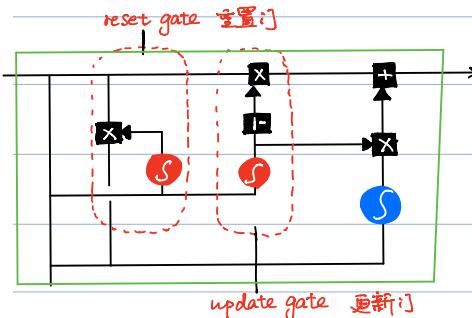
$$\tanh \in [-1, 1]$$



当向量流经神经网络的时候，由于各种数学运算，向量经历许多变换 e.g. 让一个值继续 \times 可以让这个值变成天文数字，让其他值看起来微不足道。

\tanh 确保值保持在 $-1 \sim 1$ 之间，从而调节了神经网络的输出

GRU是新一代的循环神经网络，使用隐藏状态来进行信息的传递，包含两个门：更新门和重置门



• 更新门：哪些(新)信息需要被添加

• 重置门：决定遗忘先前信息的程度

GRU计算开销小，更加适用于大规模数据集