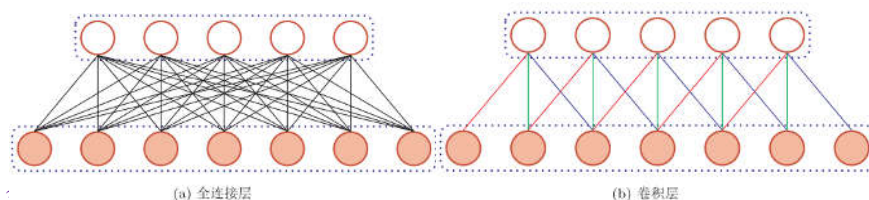
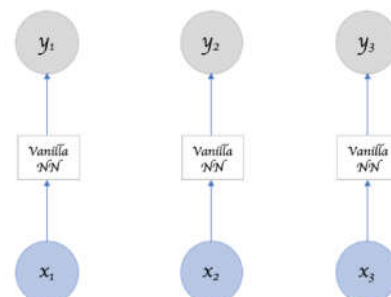


循环神经网络



前馈神经网络中的一些不足

- ◆ 连接存在层与层之间，每层的节点之间是无连接。（无循环）
 - 割裂了时序数据中相邻时间步的前后时序关系
- ◆ 输入和输出的维数都是固定的，不能任意改变。
 - 时序数据的长度通常是变化的，无法处理变长的序列数据
 - 固定长度：信息噪声（补长），信息丢失（去尾）
- ◆ 假设每次输入都是独立的
 - 每次网络的输出只依赖于当前的输入
 - 很多现实任务中，网络的输出不仅和当前时刻的输入相关，也和其过去一段时间的输出相关。比如：有限状态自动机，图灵机。

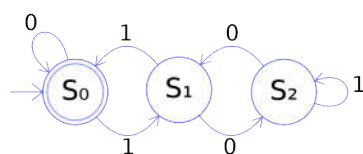


循环神经网络



确定有限自动机

- ◆ 如何用前馈神经网络模拟一个有限状态自动机？ -> 时序机



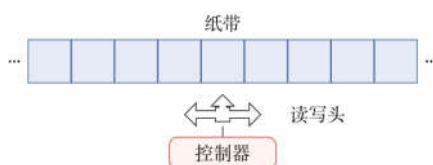
确定有限状态自动机 A 是由

- 一个非空有限的状态集合 Q
- 一个输入字母表 Σ (非空有限的字符集合)
- 一个转移函数 $\delta: Q \times \Sigma \rightarrow Q$ (例如: $\delta(q, \sigma) = p, (p, q \in Q, \sigma \in \Sigma)$)
- 一个开始状态 $s \in Q$
- 一个接受状态的集合 $F \subseteq Q$

图灵机

所组成的5-元组。因此一个DFA可以写成这样的形式: $A = (Q, \Sigma, \delta, s, F)$ 。

- ◆ 英国数学家艾伦·图灵于1936年提出，一种将人的计算行为抽象化的数学逻辑机
- ◆ 一种抽象数学模型，可以用来模拟任何可计算问题。

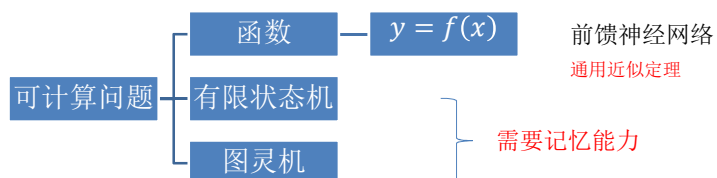


图灵的基本思想是用机器来模拟人们用纸笔进行数学运算的过程，他把这样的过程看作下列两种简单的动作：

- 在纸上写上或擦除某个符号；
- 把注意力从纸的一处移动到另一处；



循环神经网络



如何给网络增加记忆能力?

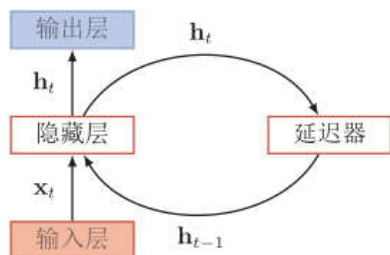
39



循环神经网络



- ◎ 循环神经网络
 - ◆ 通过使用带自反馈的神经元，能够处理任意长度的序列。
- ◎ 一般结构
 - ◆ 输入层、隐藏层和输出层
 - ◆ 前一个时刻的输出结果会被送入到下一个时刻作为输入
 - ◆ 当前时刻的输出时会综合考虑前面所有时刻的数据输入



$$h_t = \begin{cases} 0 & t = 0 \\ f(h_{t-1}, x_t) & \text{otherwise} \end{cases}$$

40

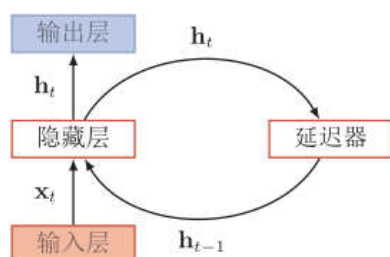


循环神经网络



循环神经网络

- 比前馈神经网络更加符合生物神经网络的结构。
- 已经被广泛应用在语音识别、语言模型以及自然语言生成等任务上。



$$h_t = \begin{cases} 0 & t = 0 \\ f(h_{t-1}, x_t) & \text{otherwise} \end{cases}$$

41



循环神经网络

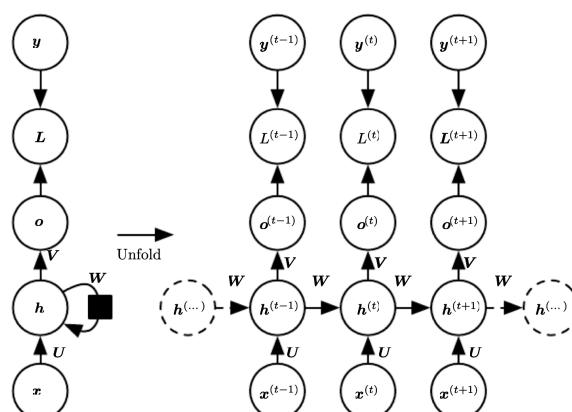


简单循环网络

- $x = \{x^0, x^1, \dots, x^{t-1}, x^t, x^{t+1}, \dots\}$: 输入, 时序性数据
- x^t : t 时刻的输入向量
- h^t : t 时刻隐状态向量
- U, V, W : 输入、隐状态和输出的权重向量

特点:

- 权重共享 (U, V, W)
 - 参数数量并不会随着时间的延伸而增加
- 隐状态 (h) 不断更新
 - 当前时刻的输出时会综合考虑前面所有时刻的数据输入



42



循环神经网络



前向传播问题

- ◆ 时序较长时，隐状态 h_t 难以表示前面所有时序的信息
- ◆ 记忆丢失

反向传播问题

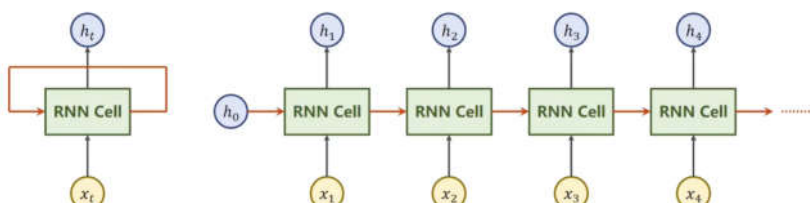
- ◆ 链式法则计算UWV梯度时，由于乘积效应，导致梯度值接近于零或者无穷大
- ◆ 梯度消失或梯度爆炸

局限性

- ◆ 简单的RNN仅适用于时序较短的模型

如何改进？

- ◆ 解决记忆丢失问题
- ◆ 解决梯度爆炸/消失问题



43

循环神经网络



长短期记忆神经网络 (LSTM)

门控单元

- ◆ 引入三个门控单元，实现记忆的选择性更新和遗忘机制
- ◆ 输入门(input)、遗忘门(forget)、输出门(output)，激活函数

长短期记忆

- ◆ 长时记忆：隐状态C
- ◆ 短时记忆：隐状态h

$$\mathbf{i}_t = \sigma(W_i \mathbf{x}_t + U_i \mathbf{h}_{t-1} + \mathbf{b}_i),$$

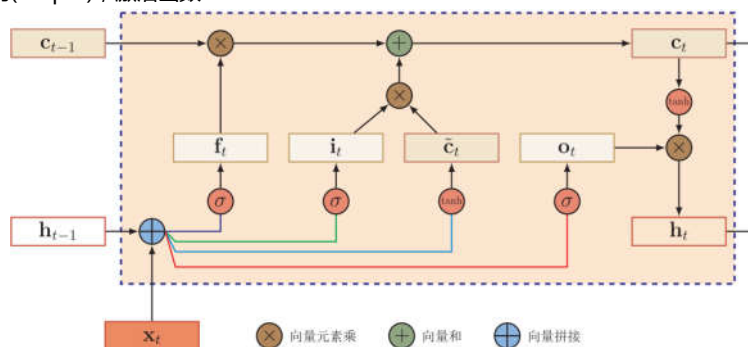
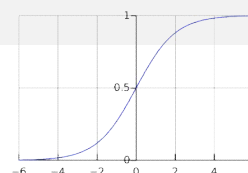
$$\mathbf{f}_t = \sigma(W_f \mathbf{x}_t + U_f \mathbf{h}_{t-1} + \mathbf{b}_f),$$

$$\mathbf{o}_t = \sigma(W_o \mathbf{x}_t + U_o \mathbf{h}_{t-1} + \mathbf{b}_o),$$

$$\tilde{\mathbf{c}}_t = \tanh(W_c \mathbf{x}_t + U_c \mathbf{h}_{t-1} + \mathbf{b}_c)$$

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t,$$

$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t),$$



44

循环神经网络 - 长短时记忆神经网络 (LSTM)



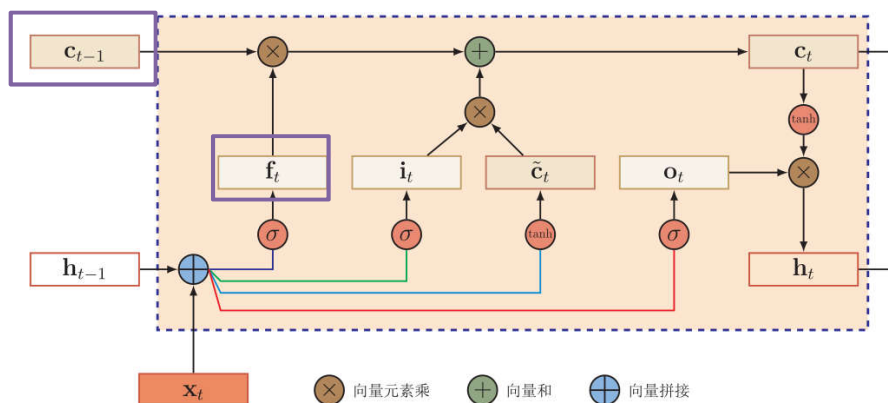
遗忘门(forget)

- 选择当前长时记忆中的哪些信息应该被延续下去，即使得长时记忆遗忘部分信息。

$$\tilde{c}_t = \tanh(W_c \mathbf{x}_t + U_c \mathbf{h}_{t-1} + \mathbf{b}_c)$$

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t,$$

$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t),$$



45

循环神经网络 - 长短时记忆神经网络 (LSTM)



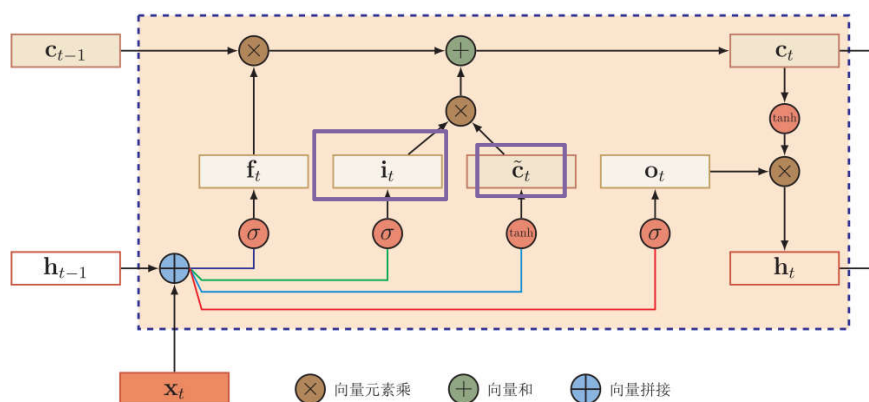
输入门

- 选择短时记忆和当前输入中的哪些信息应该被加入到长时记忆中。

$$\tilde{c}_t = \tanh(W_c \mathbf{x}_t + U_c \mathbf{h}_{t-1} + \mathbf{b}_c)$$

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t,$$

$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t),$$



46

循环神经网络 - 长短时记忆神经网络 (LSTM)



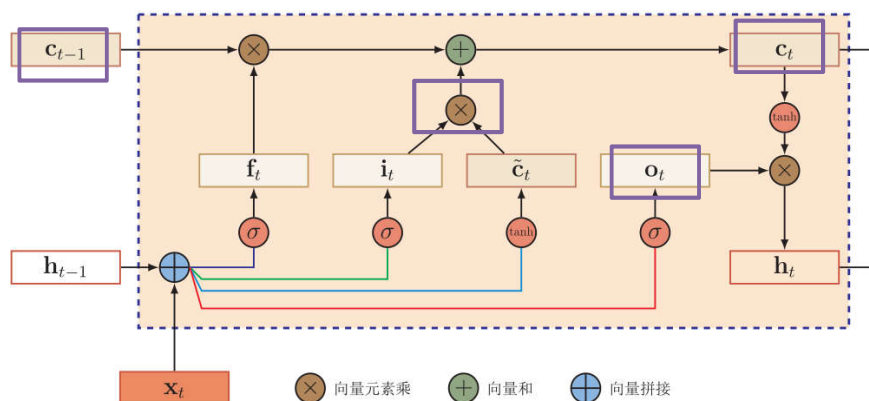
输出门

- 选择叠加当前输入信息后的长时记忆中哪些信息可以作为当前时刻的输出。

$$\tilde{c}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t$$

$$h_t = o_t \odot \tanh(c_t)$$



47

循环神经网络 - 门控循环单元网络 (GRU)



门控循环单元网络 (GRU)

- 参考LSTM，采用门控单元实现记忆的遗忘和更新机制

两个门控单元

- 在LSTM中，输入门和遗忘门其实是互补关系
- GRU将输入门和遗忘门合并为一个门：更新门，同时将输出门改进为重置门。

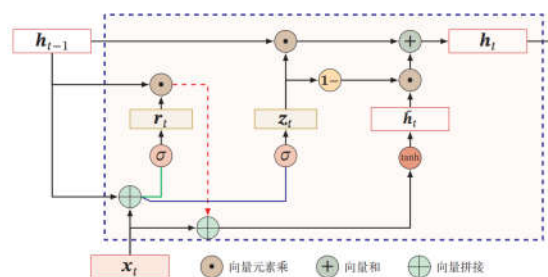


图 6.8 GRU 网络的循环单元结构

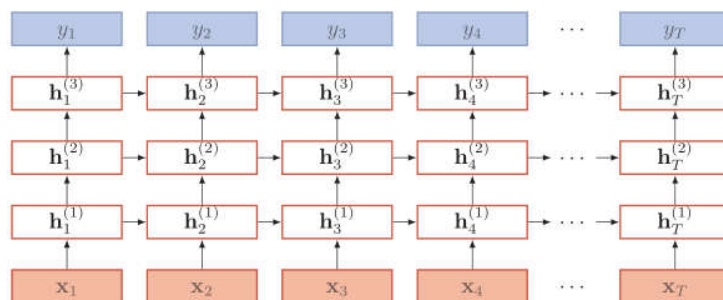
48

循环神经网络



多层循环神经网络

- 较少使用（时间开销大 & 没有显著增大“感受野” -> Transformer等新型网络）



49

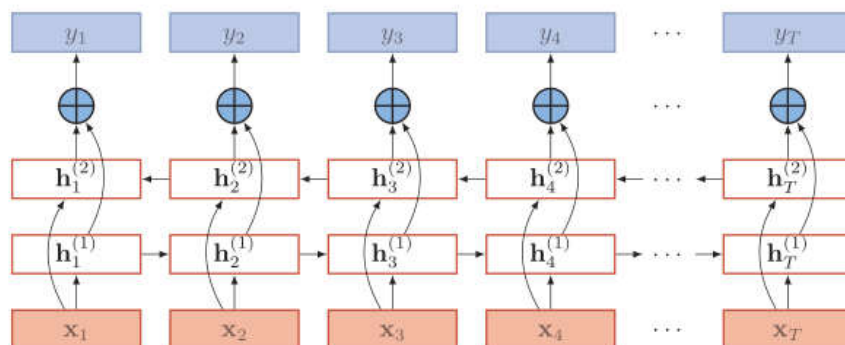


循环神经网络



双向循环神经网络

- 单向结构：仅考虑了过去输入对当前输入的作用，忽略了未来输入的作用
- 双向结构：同时考虑过去和未来的输入信息
- 双向结构不适用于需要实时预测的场景，比如实时翻译



50



循环神经网络



循环神经网络的缺陷

- ◆ 由于时间的单向流动性，多层RNN无论在层内还是层间均只能串行计算，运行开销显著增加
- ◆ 多层RNN连接并不会带来显著的效果增益

解决

- ◆ 基于位置编码和注意力策略
- ◆ 希望模型能够在学习到数据中时序位置信息同时，不丢失并行计算的优势

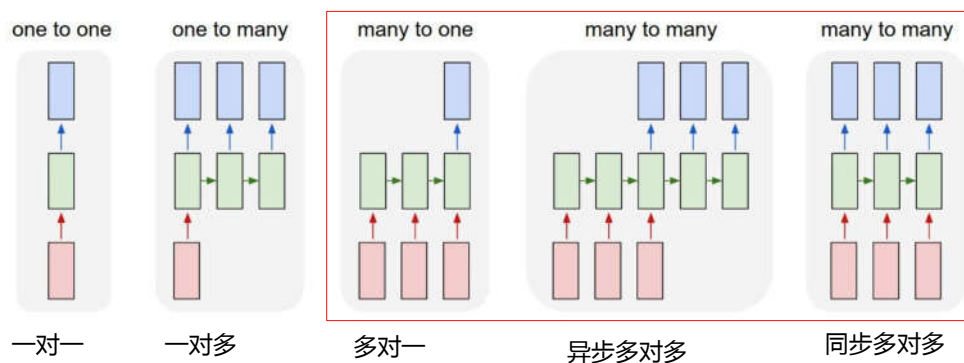
51



循环神经网络的应用



多种用法



52



循环神经网络的应用



多对一

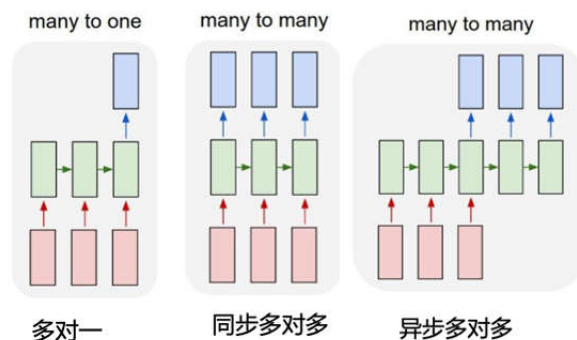
- 以多个时间步数据作为输入，输出为一个
- 各种以时序数据作为输入的分类任务

同步多对多

- 输入和输出都是多个时间步
- 同步指输出信息是实时产生的，没有滞后性
- 实时的生成式任务场景

异步多对多

- 先对输入数据进行整体建模，得到输入数据的全局表示
- 再利用此表示作为后续任务的初始隐状态，得到多个时间步的输出
- 序列到序列 (Seq2Seq)，机器翻译等对实时性要求不高的生成式场景下



53



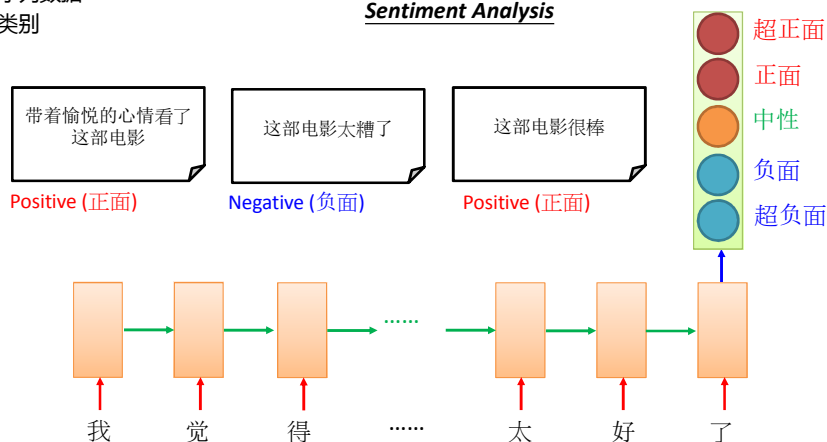
循环神经网络的应用



分类

- 输入：序列数据
- 输出：类别

Sentiment Analysis



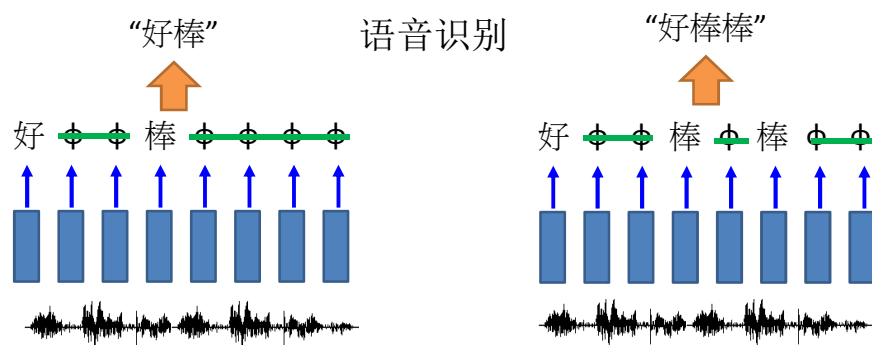
54



循环神经网络的应用



语音识别



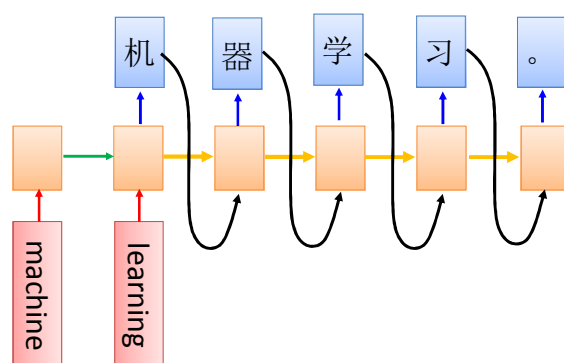
55



循环神经网络的应用



机器翻译



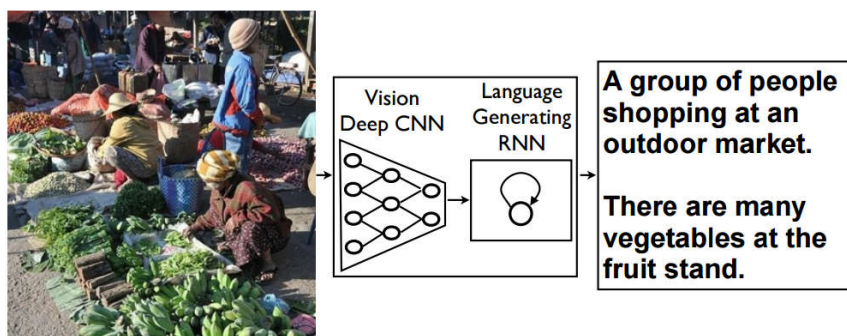
56



循环神经网络的应用



看图说话 (文本生成)



57



循环神经网络的应用



看图说话 (VQA)



Figure 5. A selection of evaluation results, grouped by human rating.

58



循环神经网络的应用



作诗

<p>白鹭窥鱼立, Egrets stood, peeping fishes. 青山照水开. Water was still, reflecting mountains. 夜来风不动, The wind went down by nightfall, 明月见楼台. as the moon came up by the tower.</p>	<p>满怀风月一枝春, Budding branches are full of romance. 未见梅花亦可人. Plum blossoms are invisible but adorable. 不为东风无此客, With the east wind comes Spring. 世间何处是前身. Where on earth do I come from?</p>
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

59



目录



- 前馈神经网络
- 卷积神经网络
- 循环神经网络
- 如何实现

60

