

Recurrent Neural Network(RNN)

简介

RNN是一种能够处理序列数据的神经网络

特点

优点

1. 能够处理变长序列
2. 能够捕捉长期依赖关系
3. 能够处理序列数据

缺点

image-20250420134243810

Long Short-term Memory(LSTM)

简介

LSTM是一种特殊的RNN

特点

1. 细胞状态 (Cell State)
 - 用于存储长期信息
 - C_t
2. 遗忘门 (Forget Gate)
 - 决定是否遗忘信息
 - f_t
3. 输入门 (Input Gate)
 - 决定是否存储新信息
 - i_t
4. 输出门 (Output Gate)
 - 决定是否输出信息
 - o_t
5. 隐藏状态 (Hidden State)
 - 用于存储短期信息
 - h_t

公式

$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$

$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$

1. 遗忘门

$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$

输入门

- W_f 遗忘门权重
- W_i 输入门权重

2. 隐藏层

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad \tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

那么

- W_i 和 W_C 是权重矩阵
- \tilde{C}_t 是候选细胞状态
- i_t 是遗忘门输出, \tilde{C}_t 是候选细胞状态, C_t 是细胞状态

3. 细胞状态

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t$$

- f_t 是遗忘门输出

4. 输出门

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

那么

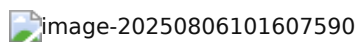
- W_o 是权重矩阵

5. 输出

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

那么

- o_t 是输出门输出, C_t 是细胞状态



那么

LSTM (长短期记忆网络) 是循环神经网络(RNN)的一种, 它通过引入细胞状态和遗忘门/输入门机制, 解决了传统RNN的梯度消失/爆炸问题。

1. **细胞状态 (Cell State)** - LSTM 的核心, 用于存储长期信息。它通过遗忘门和输入门进行更新。
2. **遗忘门** - LSTM 通过遗忘门决定丢弃哪些信息。
 - 输入: 前一个时间步的遗忘门输出和当前时间步的输入。
 - 输出: 遗忘门输出。
3. **输入门** - LSTM 通过输入门决定存储哪些信息。
 - 输入: 前一个时间步的输入门输出和当前时间步的输入。
 - 输出: 输入门输出。
4. **细胞状态更新** - LSTM 通过细胞状态更新公式更新细胞状态。
 - 输入: 前一个时间步的细胞状态、遗忘门输出、输入门输出和候选细胞状态。
 - 输出: 更新后的细胞状态。

在 LSTM 中, **Memory** 指的是细胞状态, **Input** 指的是输入门输出。

那么

GRU (门控循环单元)

GRU 是 LSTM 的简化版本, 它通过引入重置门和更新门机制, 解决了 LSTM 的复杂度问题。

那么

1. **重置门 (reset gate)**

重置门的作用是重置细胞状态, 使得模型能够捕捉到局部信息。

重置门控制重置隐藏状态 $z_t = \sigma(W_z \cdot [h_{t-1}, x_t])$

- W_z 权重矩阵
- z_t 重置门输出，范围在 [0, 1]
- z_t 控制重置隐藏状态

2. 重置门 (reset gate)

重置门控制重置隐藏状态 $r_t = \sigma(W_r \cdot [h_{t-1}, x_t])$

- W_r 权重矩阵
- r_t 重置门输出，范围在 [0, 1]

候选隐藏状态

1. 候选隐藏状态

候选隐藏状态 h_{t-1} 和重置门输出 r_t 结合 $z_t = \sigma(W_z \cdot [h_{t-1}, x_t])$

2. 候选隐藏状态

候选隐藏状态 $r_t = \sigma(W_r \cdot [h_{t-1}, x_t])$

3. 候选隐藏状态

候选隐藏状态 x_t 和重置门输出 h_{t-1} 结合 r_t 控制重置隐藏状态 $\tilde{h}_t = \tanh(W \cdot [r_t \odot h_{t-1}, x_t])$

4. 候选隐藏状态

候选隐藏状态 z_t 和重置门输出 h_{t-1} 结合 \tilde{h}_t 控制重置隐藏状态 $h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t$

Slot Filling

槽填充

1. 槽填充 (Slot Filling) 是序列标注任务中的一种
2. 槽填充 (Slot Filling) 是序列标注任务中的一种
 - 槽填充 (Slot Filling) 是序列标注任务中的一种

槽

槽填充

1. 槽填充 (Slot Filling) 是序列标注任务中的一种
2. 槽填充 (Slot Filling) 是序列标注任务中的一种
3. 槽填充 (Slot Filling) 是序列标注任务中的一种

槽填充

1. 槽填充 (Slot Filling) 是序列标注任务中的一种
2. 槽填充 (Slot Filling) 是序列标注任务中的一种
3. 槽填充 (Slot Filling) 是序列标注任务中的一种

Connectionist Temporal Classification

CTC 连接时序分类

CTC

CTC 连接时序分类 (Connectionist Temporal Classification) 是一种用于处理序列标注任务的算法

- `XX`
- `XXXXXXXXXXXX blank XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX`

11

CTC Loss “ ” loss