



人工智能与深度学习

5-2 循环神经网络 II

王中雷

经济学院、王亚南经济研究院、邹至庄经济研究院

厦门大学

(第一版)

内容摘要

1. GRU

2. LSTM

GRU

1. GRU 是 Gated Recurrent Unit (Cho et al., 2014) 的缩写
2. 重置门 (reset gate) 和更新门 (update gate) 有助于改善梯度传播, 使 GRU 更容易学习长期依赖关系
 - 引入一个候选状态的计算
 - 重置门: 控制在生成候选状态时使用多少历史信息
 - 更新门: 平衡候选状态和上一个隐藏状态, 来生成当前隐藏状态

[Cho et al. (2014), Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation, arXiv:1406.1078]

重置门 (Reset gate)

1. 重置门 (reset gate) 控制在计算候选隐藏状态时, 使用多少过去的隐藏状态信息

2. 重置门 (reset gate) 的相关计算如下:

- $\Gamma_r^{<i>} = \sigma \left\{ \mathbf{W}_r \cdot \left(\mathbf{x}^{<i>T}, \mathbf{a}^{<i-1>T} \right)^T + \mathbf{b}_r \right\} : \text{与} \mathbf{a}^{<i-1>} \text{具有相同维度}$

- $\tilde{\mathbf{a}}^{<i>} = \tanh \left\{ \mathbf{W} \cdot \left(\mathbf{x}^{<i>T}, (\Gamma_r^{<i>} \circ \mathbf{a}^{<i-1>})^T \right)^T + \mathbf{b} \right\}$

- $\Gamma_r^{<i>}$ 控制我们得到 $\tilde{\mathbf{a}}^{<i>}$ 时, 多大程度地 “忘记过去”

3. 模型参数:

- $\mathbf{W}_r, \mathbf{b}_r$ 用于计算 $\Gamma_r^{<i>}$

- $\mathbf{W}, \mathbf{b}$ 用于计算 $\tilde{\mathbf{a}}^{<i>}$

更新门 (Update gate)

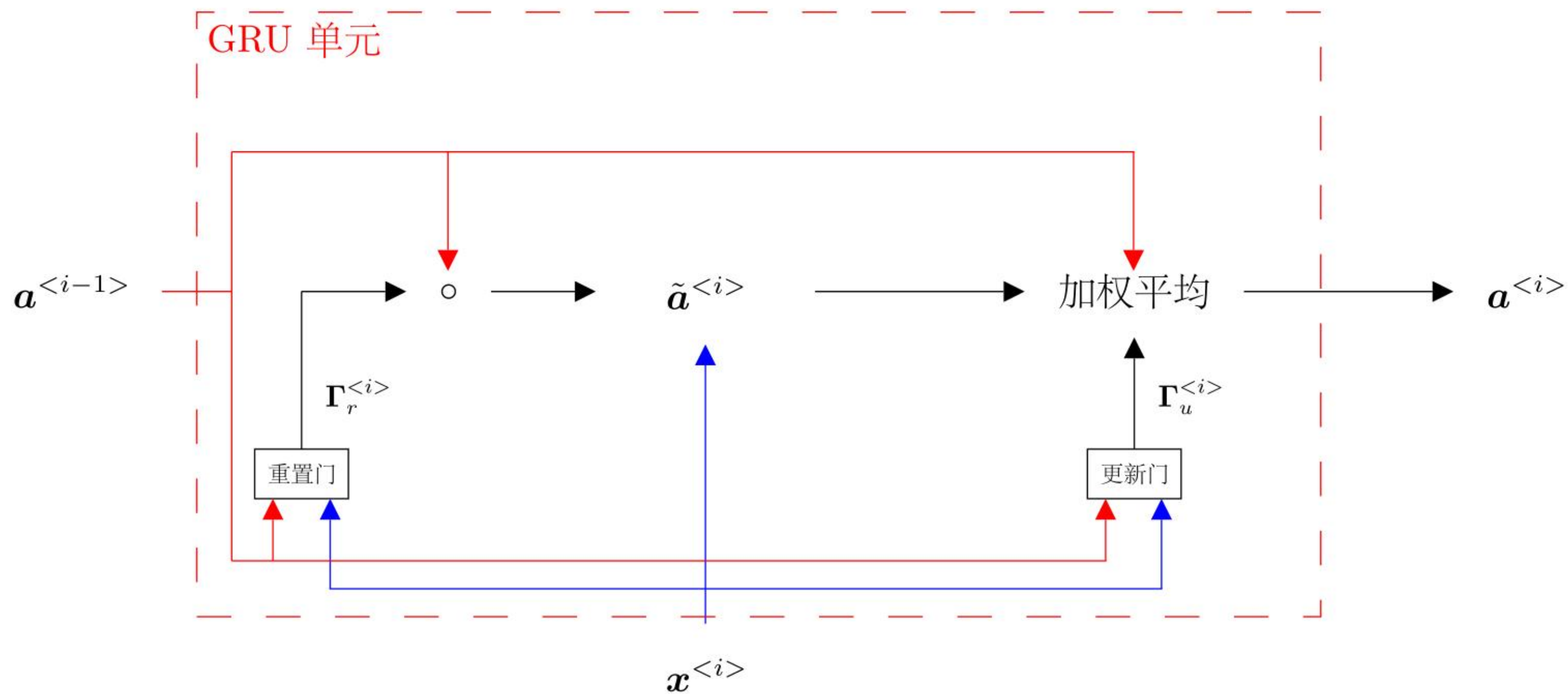
1. 更新门 (update gate) 用于得到 $\mathbf{a}^{<i>}$
2. 更新门 (update gate) 的相关计算如下:

- $\Gamma_u^{<i>} = \sigma \left\{ \mathbf{W}_u \cdot \left(\mathbf{x}^{<i>\text{T}}, \mathbf{a}^{<i-1>\text{T}} \right)^{\text{T}} + \mathbf{b}_u \right\}$: 与 $\mathbf{a}^{<i-1>}$ 具有相同维度
- $\mathbf{a}^{<i>} = (1 - \Gamma_u^{<i>}) \circ \mathbf{a}^{<i-1>} + \Gamma_u^{<i>} \circ \tilde{\mathbf{a}}^{<i>}$

3. 模型参数:

- $\mathbf{W}_u, \mathbf{b}_u$ 用于计算 $\Gamma_u^{<i>}$

流程图



LSTM

1. LSTM 是 Long Short-Term Memory unit 的缩写
2. 通过输入门 (input gate)、遗忘门 (forget gate) 以及输出门 (output gate) 等门控机制信息流，从而更好地建模不同时间尺度的依赖关系
3. 首先，产生一个候选“细胞状态” (cell state)

$$\tilde{\mathbf{c}}^{<i>} = \tanh \left\{ \mathbf{W}_c \cdot \left(\mathbf{x}^{<i>^T}, \mathbf{a}^{<i-1>^T} \right)^T + \mathbf{b}_c \right\}$$

- 其与 $\mathbf{a}^{<i-1>}$ 具有相同维度

[Hochreiter & Schmidhuber (1997), Long Short-Term Memory; 本节部分内容来自于吴恩达教授在 Coursera 上的课程 Deep Learning Specialization]

LSTM

1. 输入门 (input gate) 的相关计算如下:

$$\mathbf{\Gamma}_i^{<i>} = \sigma \left\{ \mathbf{W}_i \cdot \left(\mathbf{x}^{<i>T}, \mathbf{a}^{<i-1>T} \right)^T + \mathbf{b}_i \right\} : \text{与} \mathbf{c}^{<i-1>} \text{具有相同维度}$$

2. 遗忘门 (forget gate) 的相关计算如下:

$$\mathbf{\Gamma}_f^{<i>} = \sigma \left\{ \mathbf{W}_f \cdot \left(\mathbf{x}^{<i>T}, \mathbf{a}^{<i-1>T} \right)^T + \mathbf{b}_f \right\} : \text{与} \mathbf{c}^{<i-1>} \text{具有相同维度}$$

3. 输入门和遗忘门共同更新当前时间步的“细胞状态”

$$\mathbf{c}^{<i>} = \mathbf{\Gamma}_i^{<i>} \circ \tilde{\mathbf{c}}^{<i>} + \mathbf{\Gamma}_f^{<i>} \circ \mathbf{c}^{<i-1>}$$

4. 遗忘门控制上一时间步细胞状态中各部分信息的保留程度

LSTM

1. 输出门 (output gate) 的相关计算如下:

$$\boldsymbol{\Gamma}_o^{<i>} = \sigma \left\{ \boldsymbol{W}_o \cdot \left(\boldsymbol{x}^{<i>T}, \boldsymbol{a}^{<i-1>T} \right)^T + \boldsymbol{b}_o \right\} : \text{与} \boldsymbol{c}^{<i-1>} \text{具有相同维度}$$

2. 输出门用于得到

$$\boldsymbol{a}^{<i>} = \boldsymbol{\Gamma}_o^{<i>} \circ \tanh(\boldsymbol{c}^{<i>})$$

3. 输出门控制当前计算中的多少信息被传递到隐藏状态

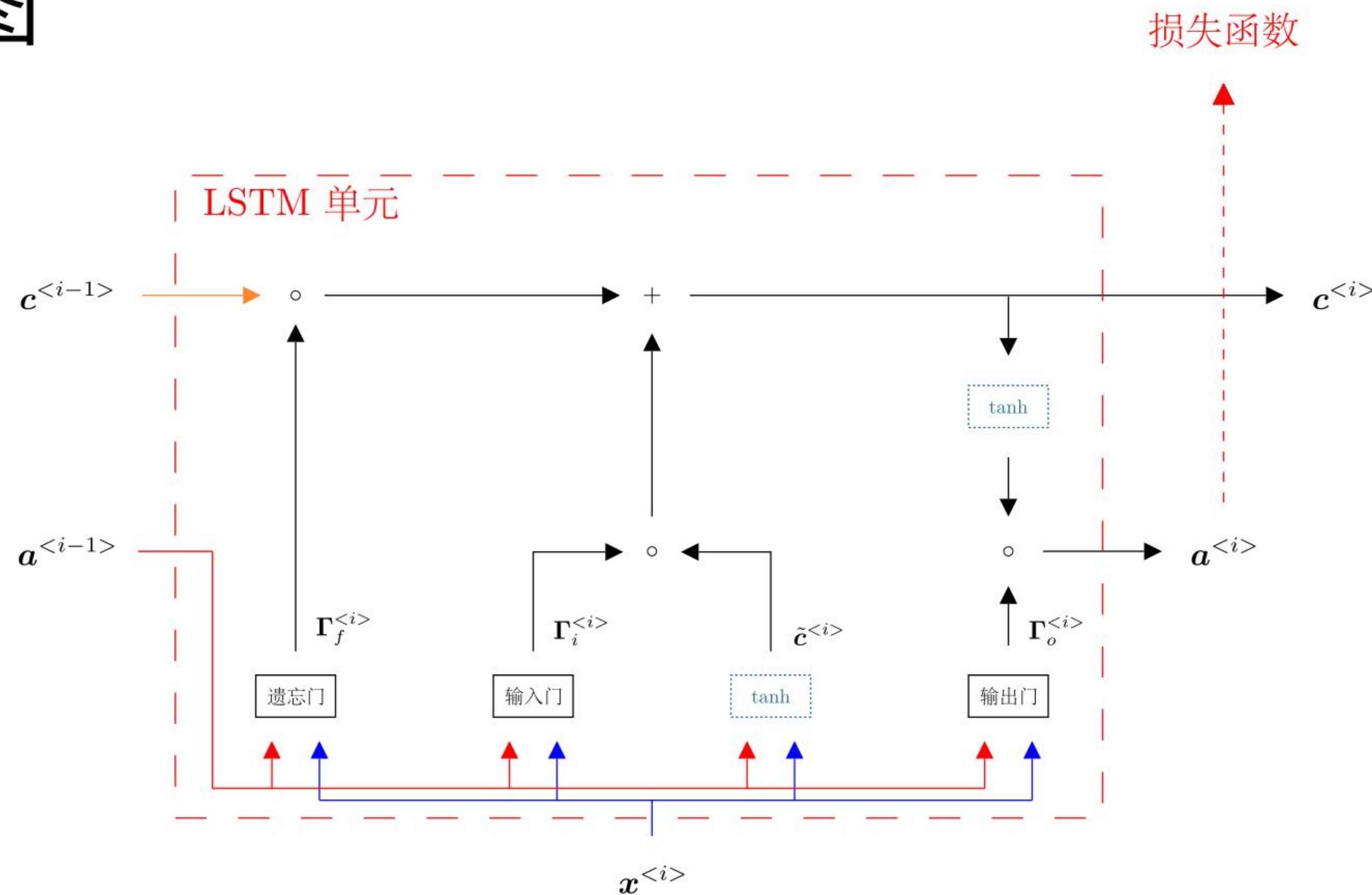
4. 初始化 $\boldsymbol{a}^{<0>} = \boldsymbol{c}^{<0>} = 0$

LSTM

1. 模型参数包括:

- W_c, b_c 用于计算候选细胞状态项
- W_i, b_i 用于输入门 (input gate) 相关计算
- W_f, b_f 用于遗忘门 (forget gate) 相关计算
- W_o, b_o 用于输出门 (output gate) 相关计算

流程图



深层循环神经网络

1. 我们可通过纵向叠加已经讨论过的“循环结构”，增加模型深度
2. 循环神经网络的两个维度
 - 时间：同一层循环单元，在不同时间步上反复应用
 - 层堆叠：同一个时间步内部，第一层循环单元的输出，继续送入第二层