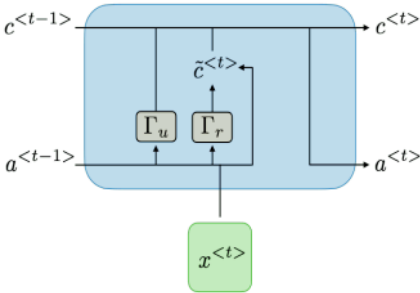
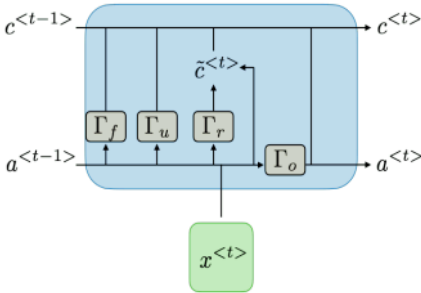


LSTM - Long Short-Term Memory

Gated Recurrent Unit (GRU) và Các đơn vị bộ nhớ dài-ngắn hạn (LSTM) đối phó với vấn đề *vanishing gradient* khi gặp phải bằng mạng RNNs truyền thống, với LSTM là sự tổng quát của GRU. Phía dưới là bảng tổng kết các phương trình đặc trưng của mỗi kiến trúc:

	Gated Recurrent Unit (GRU) (GRU)	Bộ nhớ dài-ngắn hạn (LSTM) (LSTM)
$\tilde{c}^{<t>}$	$\tanh(W_c[\Gamma_r \star a^{<t-1>}, x^{<t>}] + b_c)$	$\tanh(W_c[\Gamma_r \star a^{<t-1>}, x^{<t>}] + b_c)$
$c^{<t>}$	$\Gamma_u \star \tilde{c}^{<t>} + (1 - \Gamma_u) \star c^{<t-1>}$	$\Gamma_u \star \tilde{c}^{<t>} + \Gamma_f \star c^{<t-1>}$
$a^{<t>}$	$c^{<t>}$	$\Gamma_o \star c^{<t>}$
Các phụ thuộc		

1. Gated Recurrent Unit (GRU)

- GRU là một biến thể đơn giản hơn của LSTM, với cấu trúc giảm thiểu số lượng cổng và trọng số.
- Cấu trúc của GRU bao gồm:
 - Cổng Cập Nhật (Γ):** Quyết định thông tin nào cần được cập nhật trong trạng thái ô.
 - Cổng Nhập ($c^{<t>}$):** Sử dụng hàm tanh để tính toán giá trị đầu vào cho trạng thái ô.

2. Long Short-Term Memory (LSTM)

- LSTM phức tạp hơn GRU, với ba cổng: Cổng Quên, Cổng Nhập và Cổng Xuất.
- Các cổng này cho phép LSTM quản lý thông tin một cách linh hoạt hơn.

2.1. Các Thành Phần Chính

- Trạng thái ô (c):** Đại diện cho bộ nhớ dài hạn.
- Trạng thái ẩn (h):** Đại diện cho bộ nhớ ngắn hạn.
- Cổng (Gates):** Kiểm soát dòng thông tin:
 - Cổng Quên (f):** Quyết định thông tin nào sẽ bị bỏ qua từ trạng thái ô.
 - Cổng Nhập (i):** Quyết định thông tin mới nào sẽ được lưu trữ trong trạng thái ô.
 - Cổng Xuất (o):** Quyết định phần nào của trạng thái ô sẽ được xuất ra như trạng thái ẩn.

2.2. Hàm Kích Hoạt

- **Sigmoid (σ \sigma):** Để các cổng xuất giá trị giữa 0 và 1.
- **Tanh:** Để làm cho giá trị trạng thái ô nằm giữa -1 và 1.

Các Biến Thể LSTM

- **LSTM Hai Chiều:** Xử lý các chuỗi theo cả hai hướng tiến và lùi, cải thiện khả năng hiểu ngữ cảnh.
- **LSTM Chồng:** Nhiều lớp LSTM xếp chồng lên nhau để học sâu hơn.

Hyperparameters

- **Number of Layers:** kiểm soát depth.
- **Hidden Units:** quyết định kích thước hidden state.
- **Learning Rate:** Điều chỉnh tốc độ học của mô hình.

So Sánh

- **GRU:** Ít tham số hơn, nhanh hơn trong việc tính toán, dễ dàng hơn trong việc điều chỉnh.
- **LSTM:** Khả năng học các phụ thuộc dài hạn tốt hơn, thích hợp cho các nhiệm vụ phức tạp hơn.