

LSTM

Thành viên:

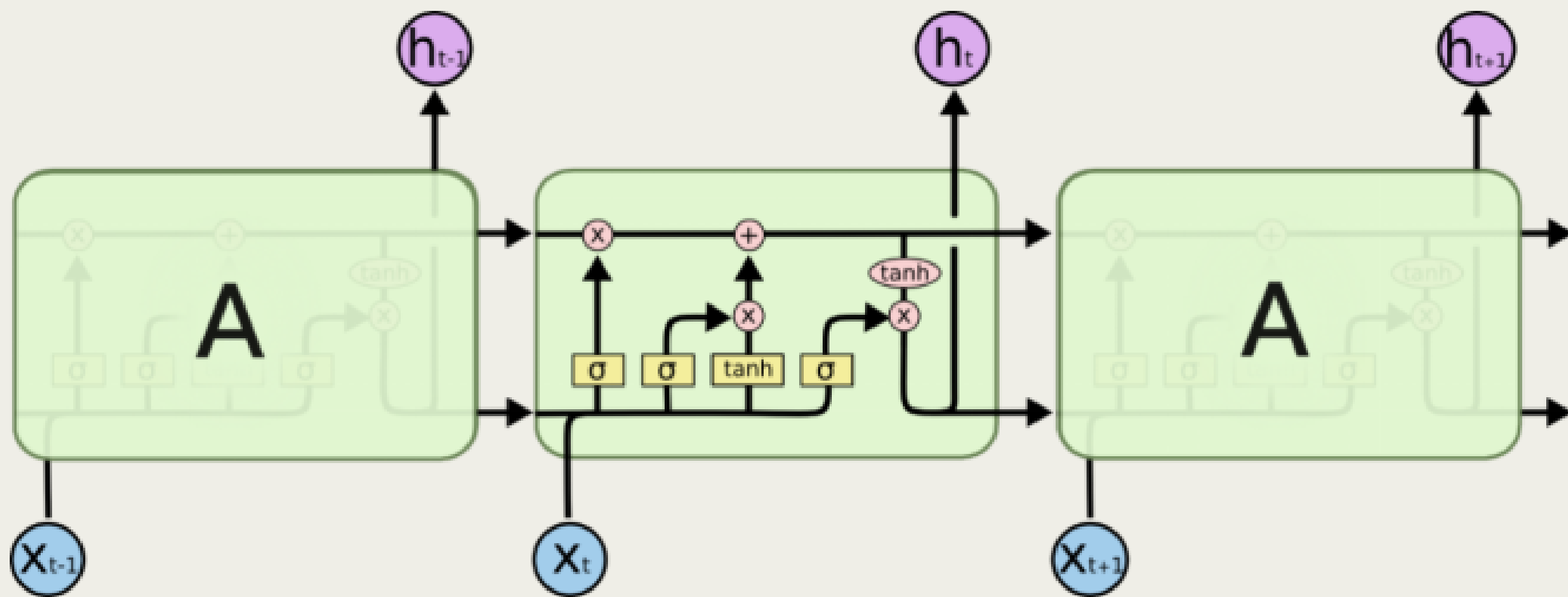
- Võ Luyện
- Trần Thị Vẹn
- Nguyễn Thanh Tú

L S T M

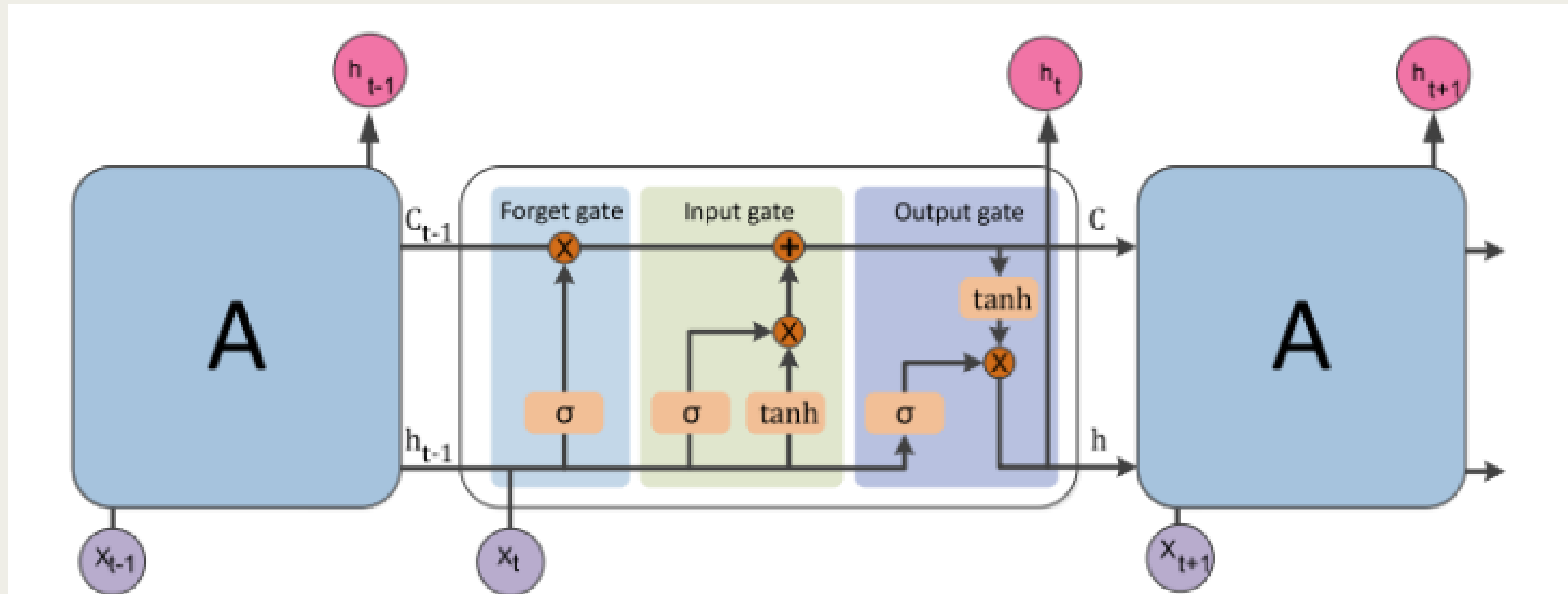
Vấn đề của RNN: Vanishing Gradient.

Xảy ra khi huấn luyện mô hình với số lượng hidden layers lớn dẫn đến việc mô hình RNN không học được các thông tin ở xa trong quá khứ.

Cần một mô hình sử dụng Long-term Memory để mô hình có thể học và lưu trữ. => Long short term memory (LSTM) ra đời.



Mô hình LSTM

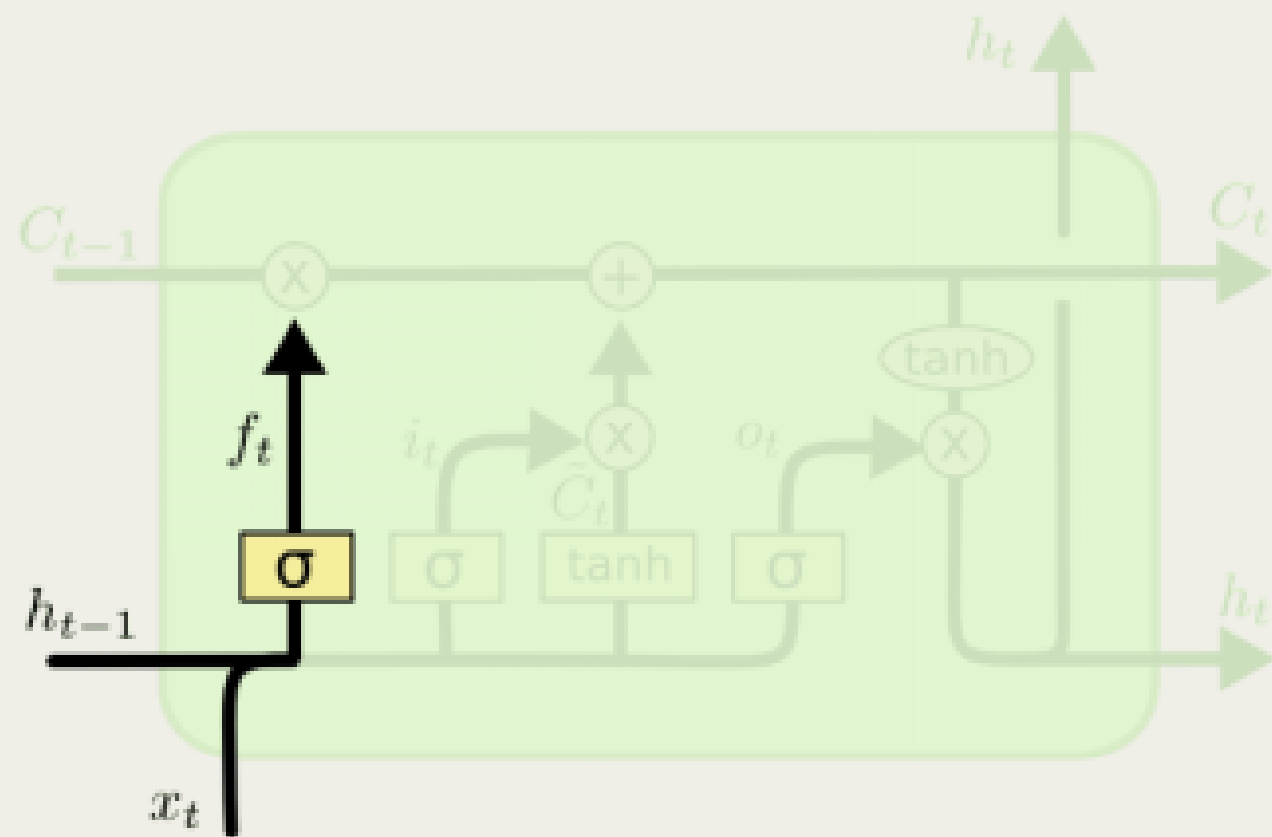


Cấu trúc của mô hình LSTM có 3 cổng:

- Forget gate: Đưa ra quyết định sẽ loại bỏ thông tin nào từ cell state
- Input gate: Quyết định thông tin nào sẽ được cập nhật vào cell state
- Output gate: Quyết định giá trị Output

CÁCH VẬN HÀNH CỦA LSTM

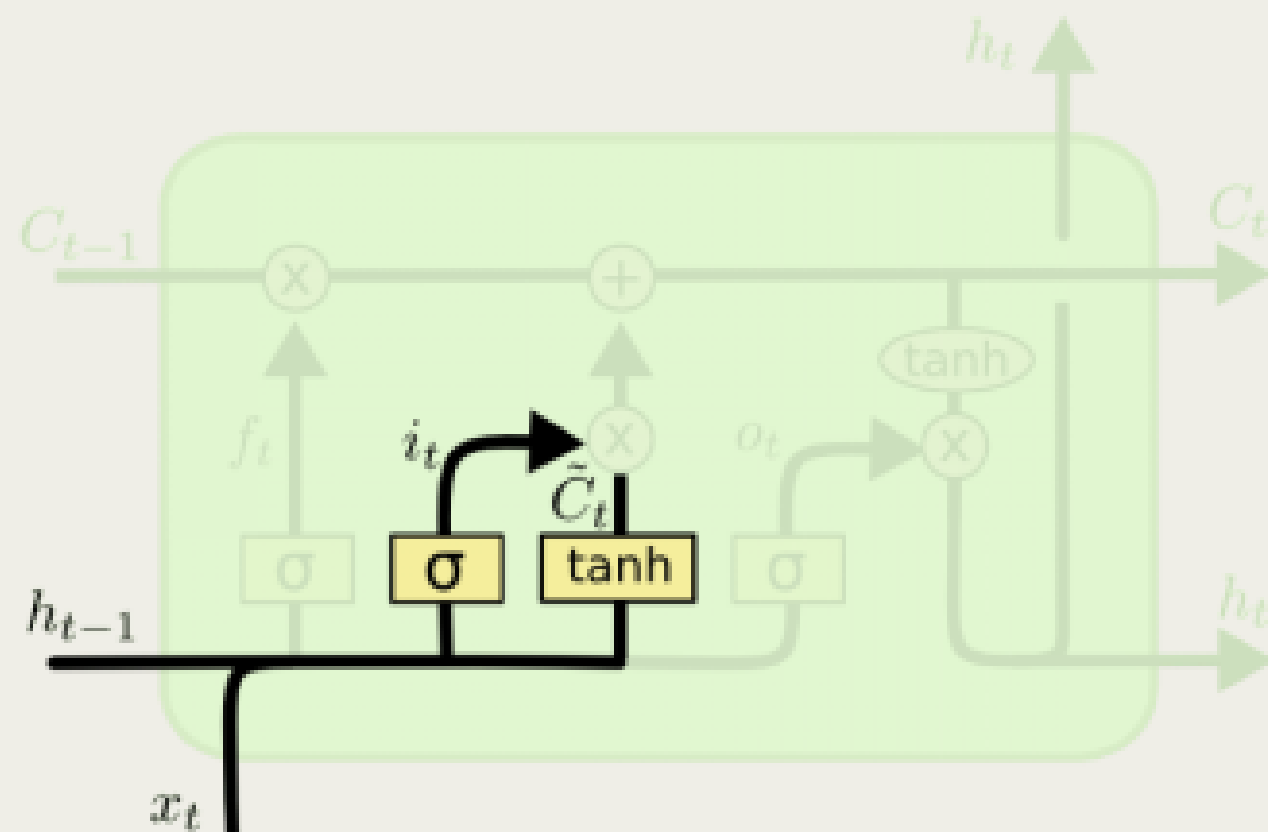
Bước 1 (Forget Gate): LSTM sẽ quyết định xem thông tin nào chúng ta sẽ cho phép đi qua ô trạng thái (cell state). Nó được kiểm soát bởi hàm sigmoid có giá trị trong khoảng $[0, 1]$. Với 1 là giữ lại toàn bộ thông tin và 0 là loại bỏ toàn bộ thông tin.



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

CÁCH VẬN HÀNH CỦA LSTM

Bước 2 (Input Gate): Quyết định thông tin mới nào được thêm vào Cell State bằng cách đưa dữ liệu qua tầng Sigmoid và tầng Tanh.

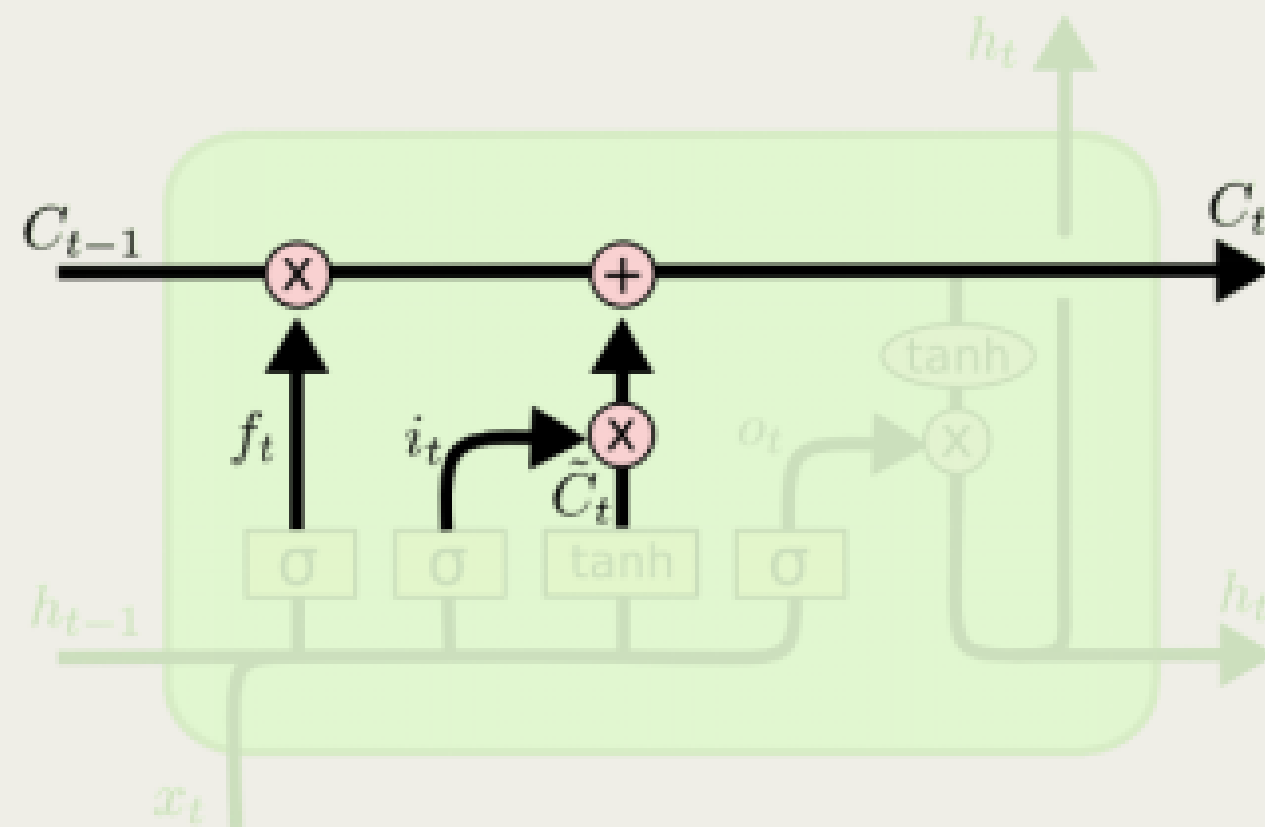


$$i_t = \sigma (W_i \cdot [h_{t-1}, x_t] + b_i)$$

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

CÁCH VẬN HÀNH CỦA LSTM

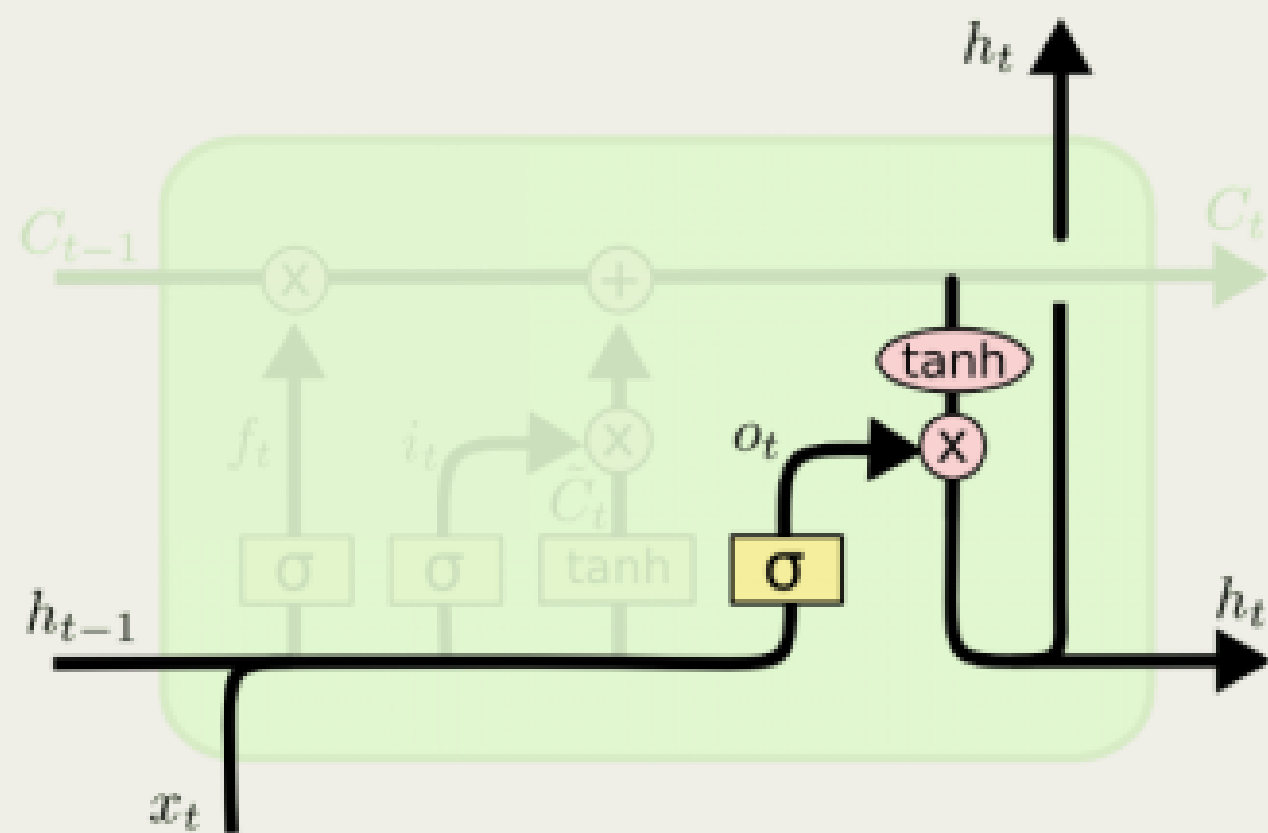
Bước 3: Cập nhật trạng thái Cell State cũ thành Cell State mới .



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

CÁCH VẬN HÀNH CỦA LSTM

Bước 4 (Output Gate): Tại đây cũng sử dụng một tầng Sigmoid để chọn ra thông tin từ Hidden State trước đó và giá trị Input hiện tại. Kết quả Output của Hidden State hiện tại là tích của Cell State sau khi đi qua hàm Tanh và kết quả của hàm Sigmoid.



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

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

Thank you!
