

# Mạng LSTM và ứng dụng trong nhận dạng thực thể – Bài 11

AI Academy Vietnam

# Thông tin giảng viên

---

- TS. Phan Việt Anh
- Email: [anhpv@lqdtu.edu.vn](mailto:anhpv@lqdtu.edu.vn)
- Mobile: 0975 639 757
- Research interest: machine learning, natural/programming language processing.

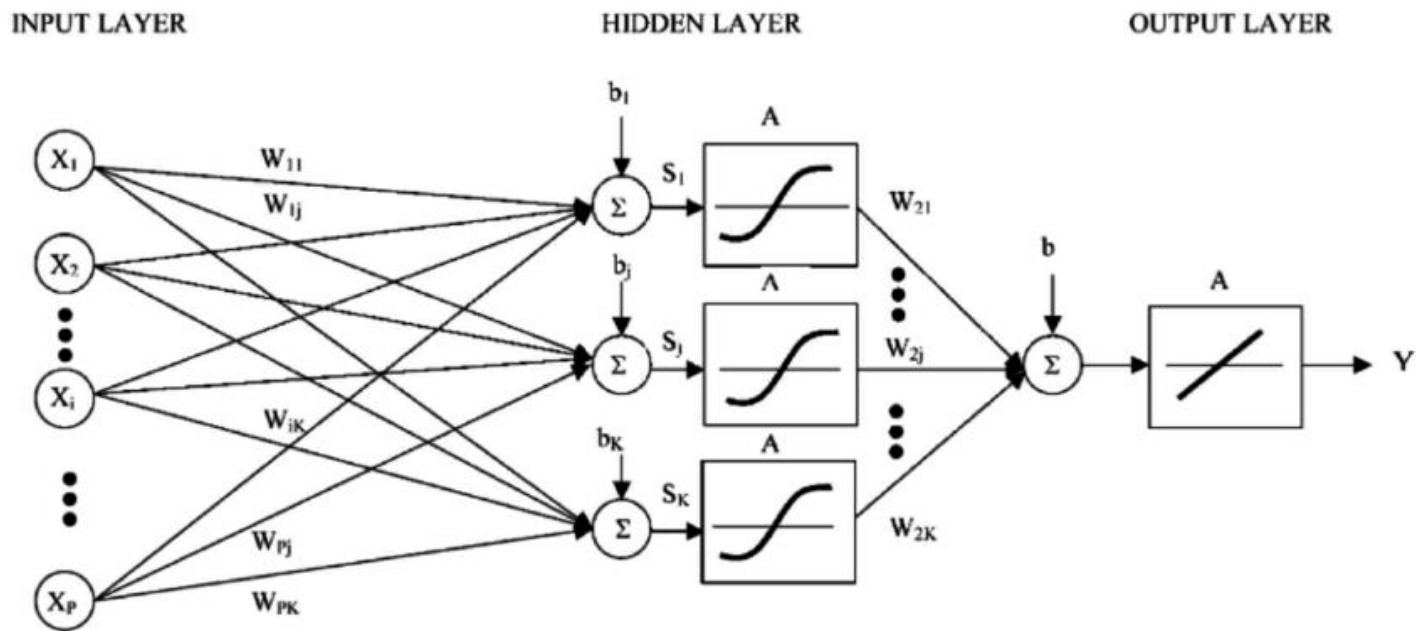
# Nội dung buổi học

---

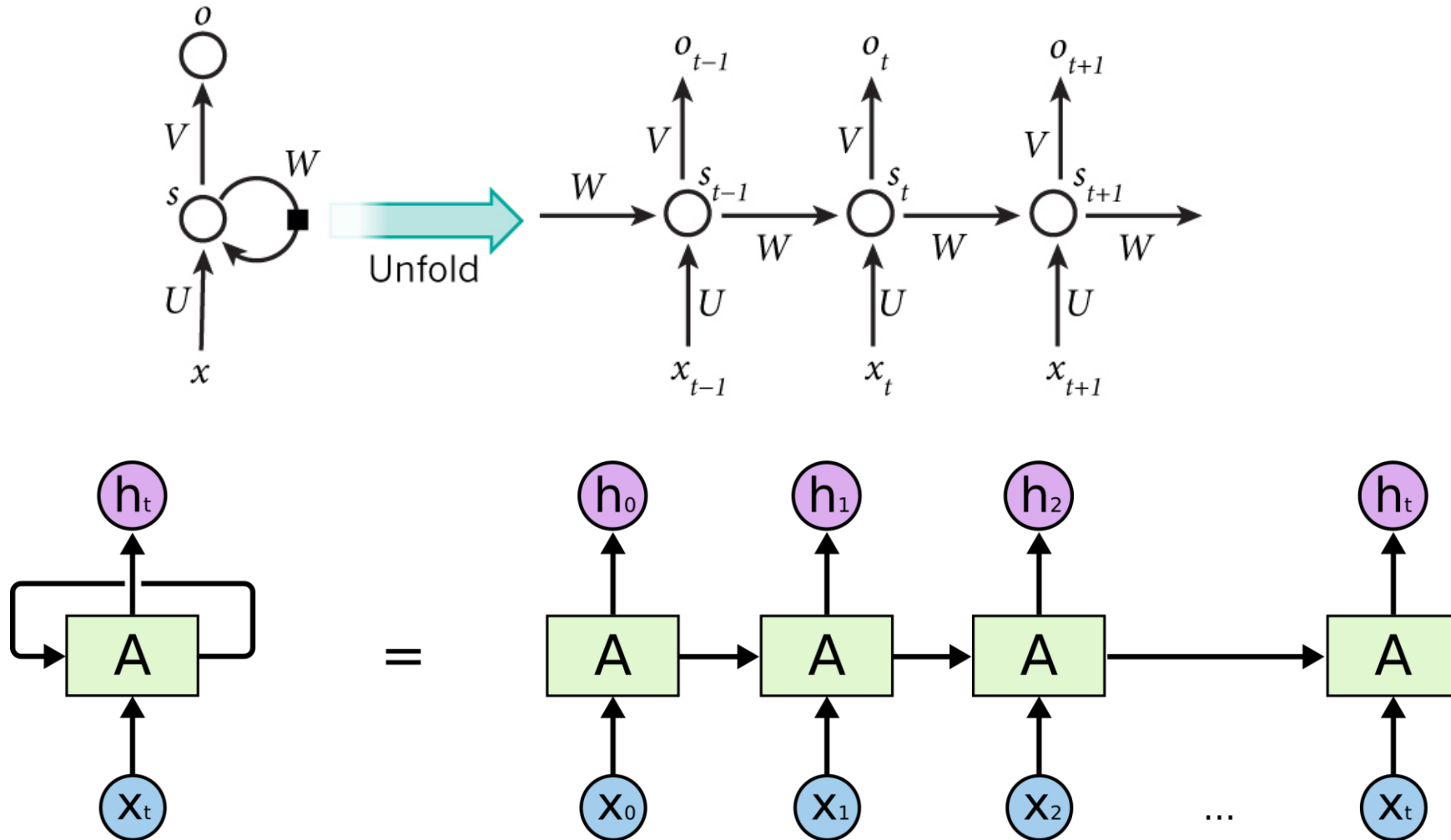
- Giới thiệu RNN
- LSTM, Bi-LSTM và ứng dụng
  - LSTM cho mô hình ngôn ngữ
  - LSTM cho phân loại
  - LSTM cho gán nhãn chuỗi
  - Sử dụng LSTM cho NER

# GIỚI THIỆU RNN

# Mạng nơ ron truyền thẳng



# Mạng nơ ron hồi quy

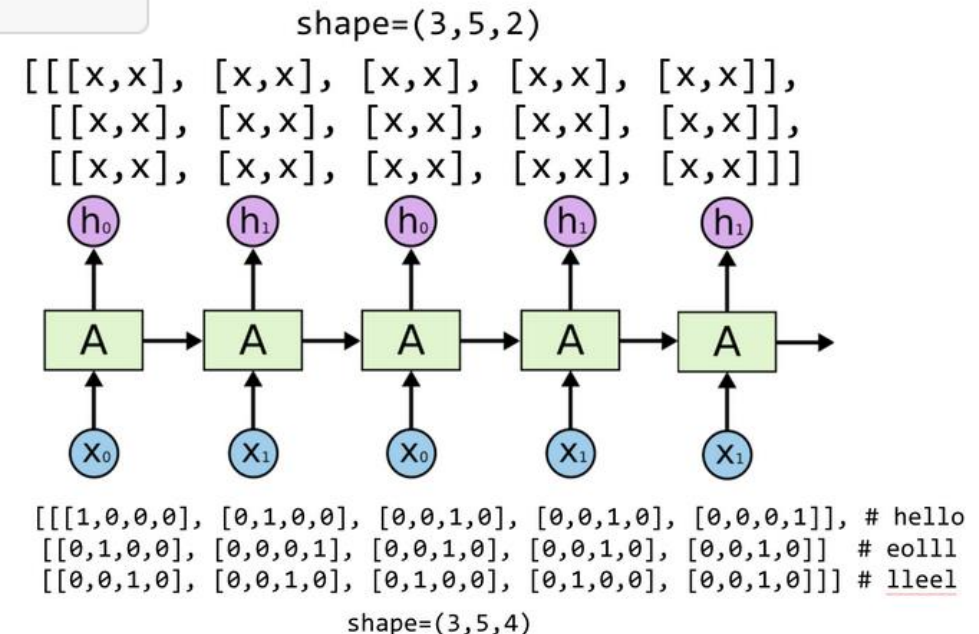


# Ví dụ một cấu trúc RNN

```
with tf.variable_scope('two_sequences') as scope:
    # One cell RNN input_dim (4) -> output_dim (2). sequence: 5
    hidden_size = 2
    cell = tf.keras.layers.SimpleRNNCell(units=hidden_size)
    x_data = np.array([[h, e, l, l, o]], dtype=np.float32)
    print(x_data.shape)
    pp.pprint(x_data)
    outputs, _states = tf.nn.dynamic_rnn(cell, x_data, dtype=tf.float32)
    sess.run(tf.global_variables_initializer())
    pp.pprint(outputs.eval())
```

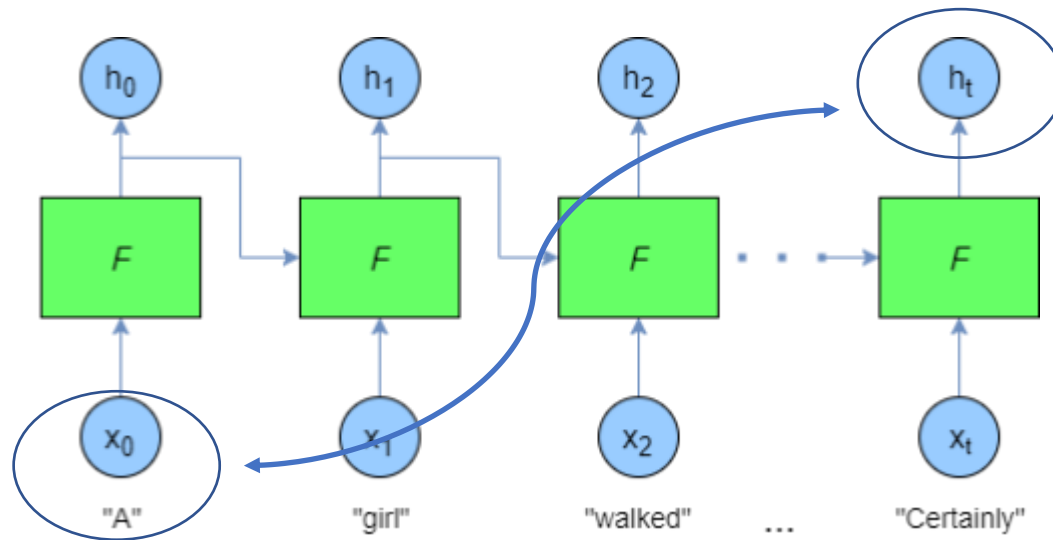
hidden\_size=2  
sequacne\_length=5  
batch = 3

# One hot encoding  
h = [1, 0, 0, 0]  
e = [0, 1, 0, 0]  
l = [0, 0, 1, 0]  
o = [0, 0, 0, 1]



# Các vấn đề với RNN

- Cấu trúc của RNN rất phù hợp với chuỗi thời gian
  - Vấn đề gặp phải: Phụ thuộc xa, nó dễ “quên thông tin cũ”
- => Mong muốn: RNN có khả năng “nhớ”



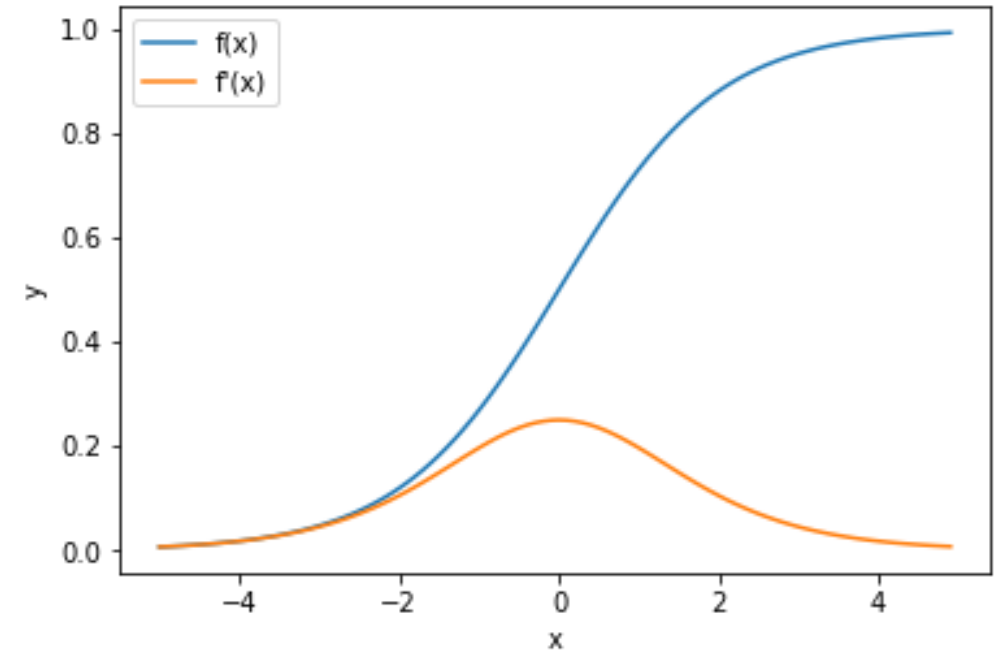


# Vanishing gradient

- Sự phụ thuộc giữa  $h_t$  và  $x_0$  có thể được mô tả bởi  $\frac{\partial h_t}{\partial x_0}$
- Nó có thể được phân tích thành

$$\frac{\partial h_t}{\partial x_0} = \frac{\partial h_t}{\partial h_{t-1}} \frac{\partial h_{t-1}}{\partial h_{t-2}} \frac{\partial h_{t-2}}{\partial h_{t-3}} \dots \frac{\partial h_0}{\partial x_0}$$

- Phụ thuộc vào hàm kích hoạt



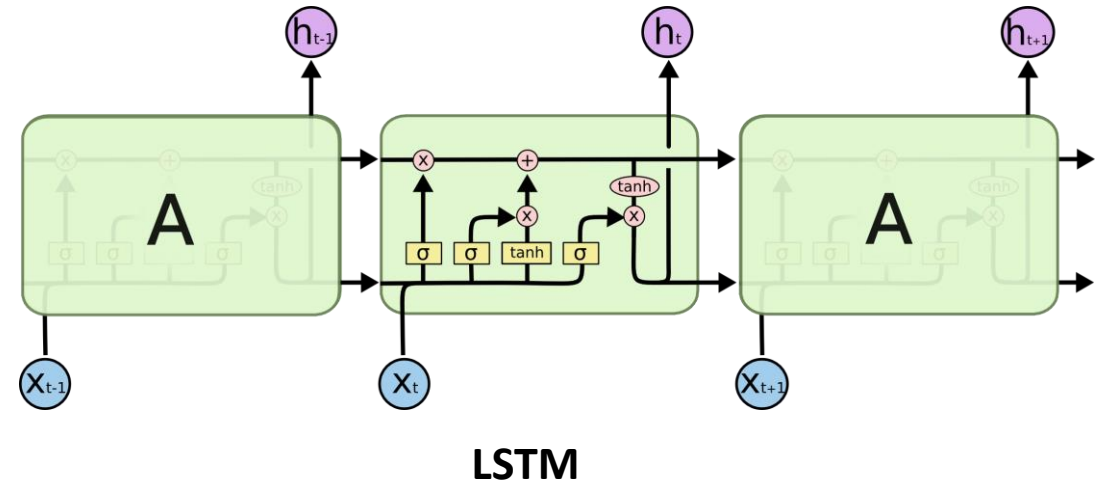
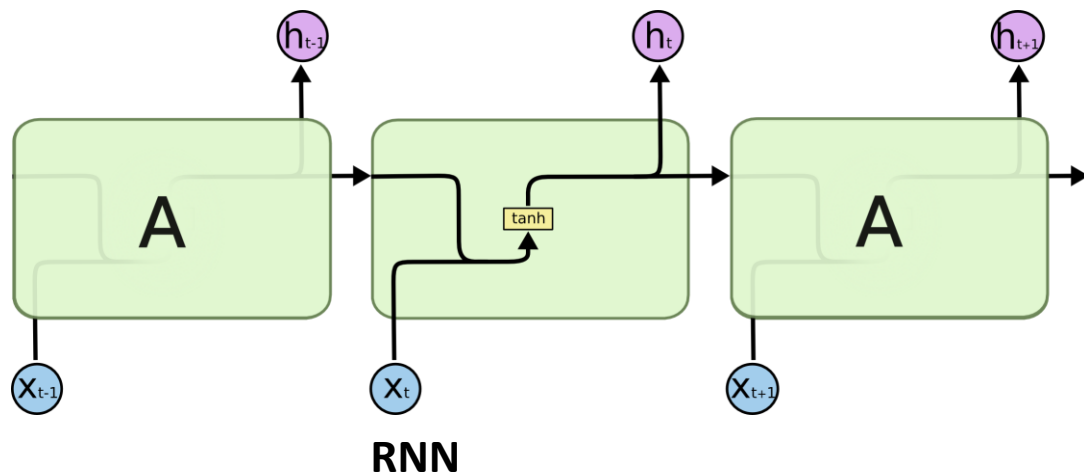
# Vanishing gradient

---

- Vấn đề gặp phải
  - Đạo hàm từng phần có thể tiệm cận 0
  - Đầu ra không liên quan đến các đầu vào cách xa phía trước
  - Sự phụ thuộc dài sẽ bị biến mất
- Giải pháp
  - Thay đổi hàm kích hoạt
    - Vấn đề có thể giảm nhưng không hoàn toàn được giải quyết
    - Có thể ảnh hưởng đến hiệu năng học
  - LSTM

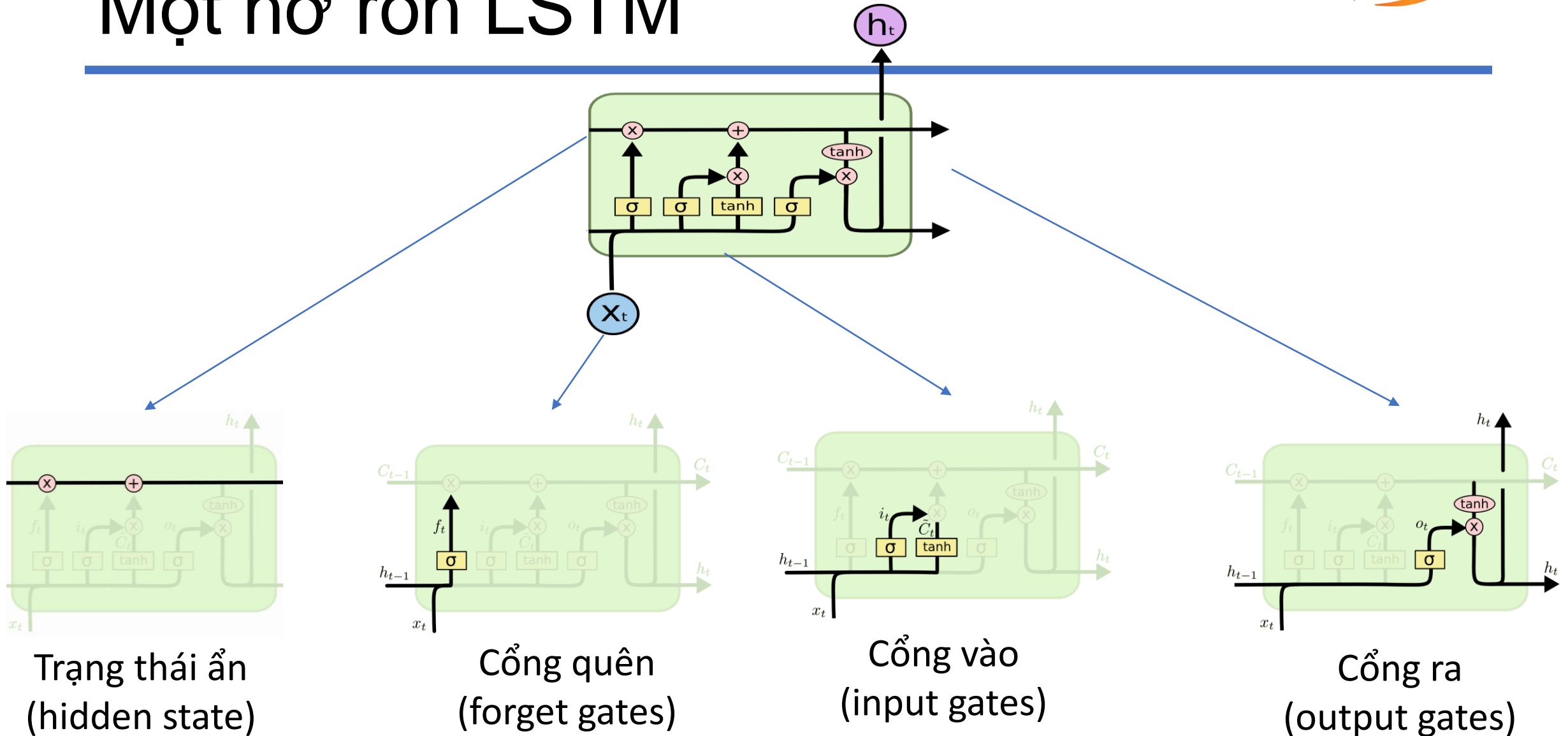
# LSTM, Bi-LSTM VÀ ỨNG DỤNG

# Long Short-Term Memory

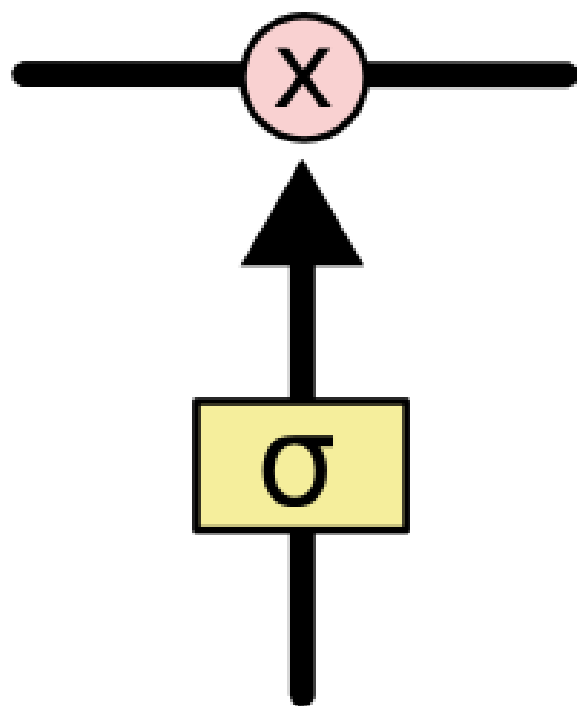


- Ý tưởng chính: Sử dụng một đường **trực tiếp** từ các **đầu vào** phía trư**ớc** đến **đầu ra hiện tại**

# Một nơ ron LSTM

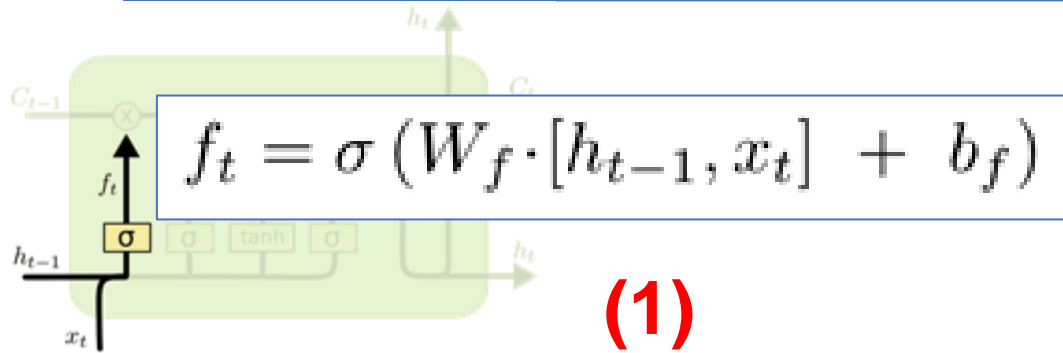


# Các cổng

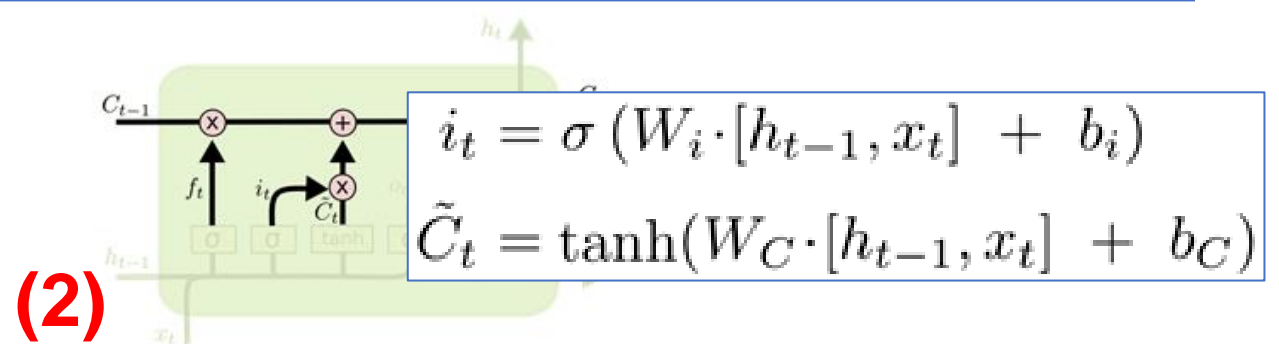


- Một nơ ron  $\sigma$
- Giá trị đầu ra  $[0, 1]$
- Nhân từng phần tử (pointwise multiplication)
- Hàm lọc
- Mục đích: Điều khiển luồng thông tin vào/ra bộ nhớ

# Tính toán trong nơ ron

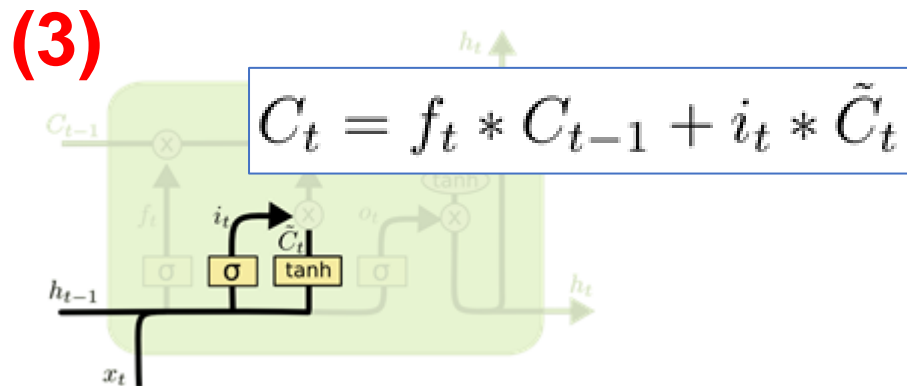


(1)



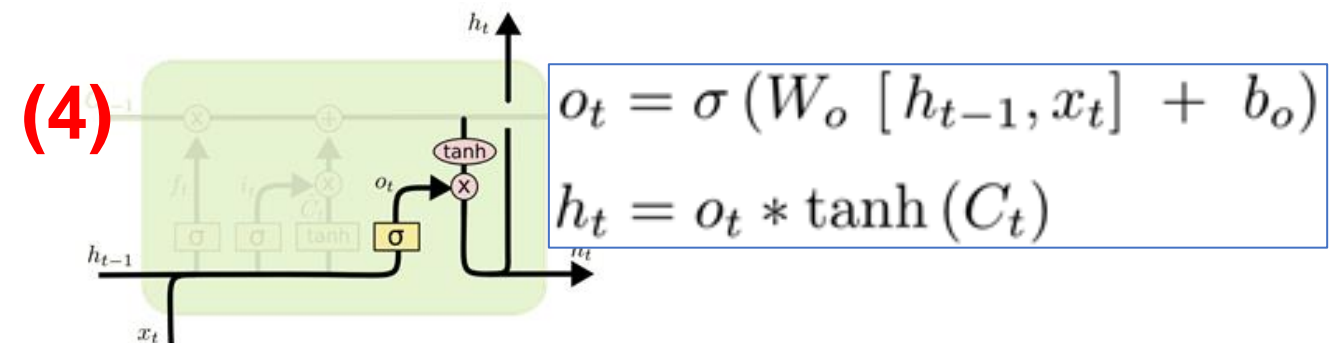
(2)

$i_t$  quyết định thành phần nào sẽ được cập nhật  $C_t$  cung cấp nội dung thay đổi



(3)

Cập nhật trạng thái nơ ron



(4)

Quyết định thành phần nào của trạng thái nơ ron sinh đầu ra

# LSTM cho mô hình ngôn ngữ



# Mô hình ngôn ngữ (language modeling)

---

- Tính xác suất của một câu
  - $P(W) = P(w_1, w_2, \dots, w_n)$
  - Ví dụ
    - $S_1 = \text{"con mèo nhảy qua con chó"} , P(S_1) \sim 1$
    - $S_2 = \text{"qua con mèo con chó nhảy"} , P(S_2) \sim 0$
- Dự đoán xác suất của từ tiếp theo
  - $P(w_5 | w_1, w_2, w_3, w_4)$
- Mô hình ngôn ngữ: Là mô hình tính
  - $P(W)$
  - hoặc  $P(w_n | w_1, w_2, \dots, w_{n-1})$

# Ứng dụng của mô hình ngôn ngữ

---

- Dịch máy
  - $P(\text{high winds tonight}) > P(\text{large winds tonight})$
- Sửa lỗi văn bản
  - The office is about fifteen **minuets** from my house
  - $P(\text{"about fifteen minutes from"}) > P(\text{about fifteen minuets from})$
- Nhận dạng giọng nói
  - $P(\text{I saw a van}) > P(\text{eyes awe of an})$
- Tóm tắt, hỏi – đáp, ..

# RNN xác định xác xuất câu

output distribution

$$\hat{y} = \text{softmax}(W_2 h^{(t)} + b_2)$$

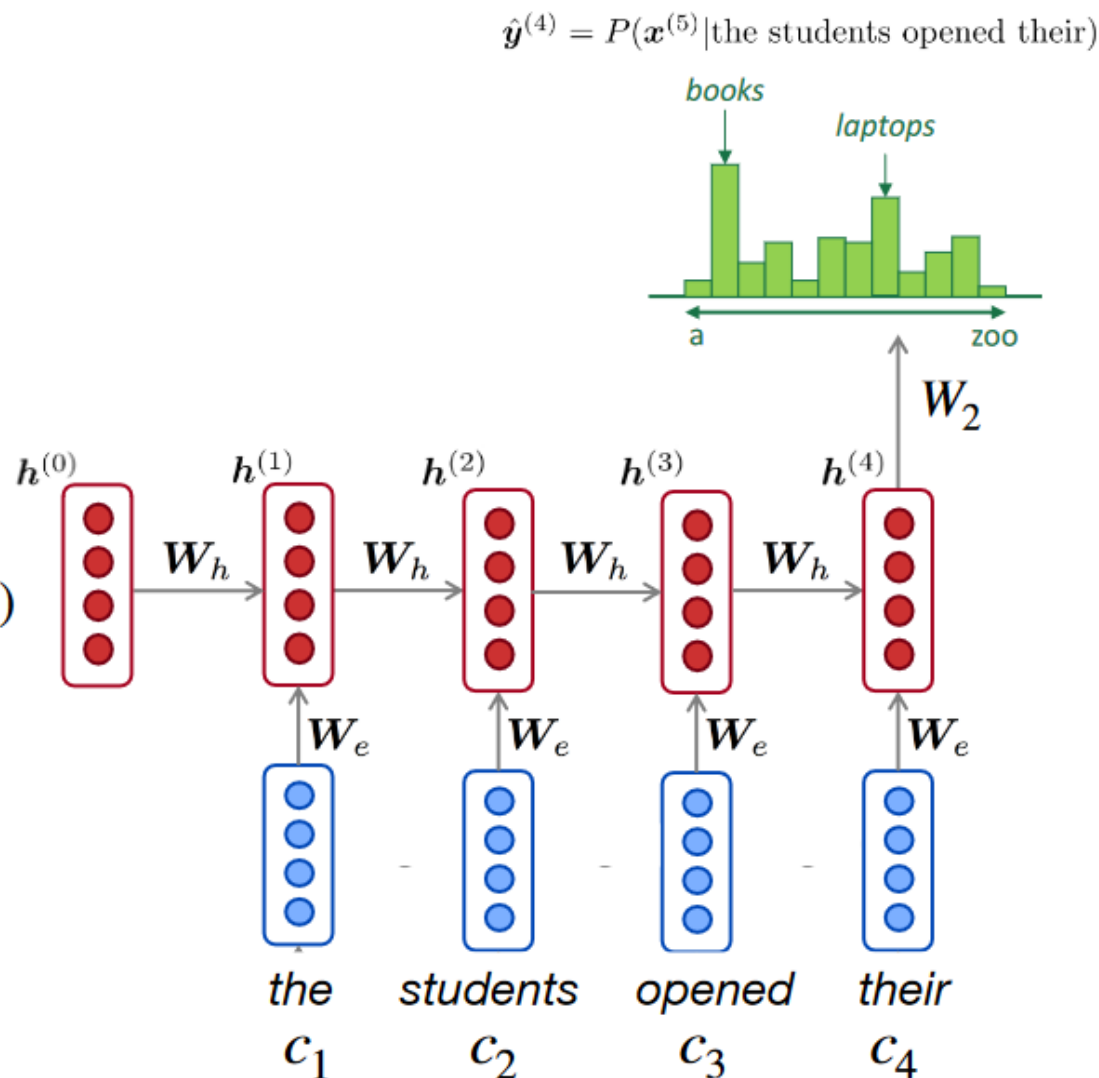
hidden states

$$h^{(t)} = f(W_h h^{(t-1)} + W_e c_t + b_1)$$

$h^{(0)}$  is initial hidden state!

word embeddings

$c_1, c_2, c_3, c_4$



# LSTM cho dịch máy

---

- Dịch: Chuyển từ câu nguồn  $f$  (tiếng Pháp) sang câu đích  $e$  (tiếng Anh)
- Ước lượng  $P(e|f)$  sử dụng NN

$$\begin{aligned} p(e|f) &= p(e_1, e_2, \dots, e_m | f) \\ &= p(e_1 | f) \cdot p(e_2 | e_1, f) \cdot p(e_3 | e_2, e_1, f) \cdot \dots \\ &= \prod_{i=1}^m p(e_i | e_1, \dots, e_{i-1}, f) \end{aligned}$$

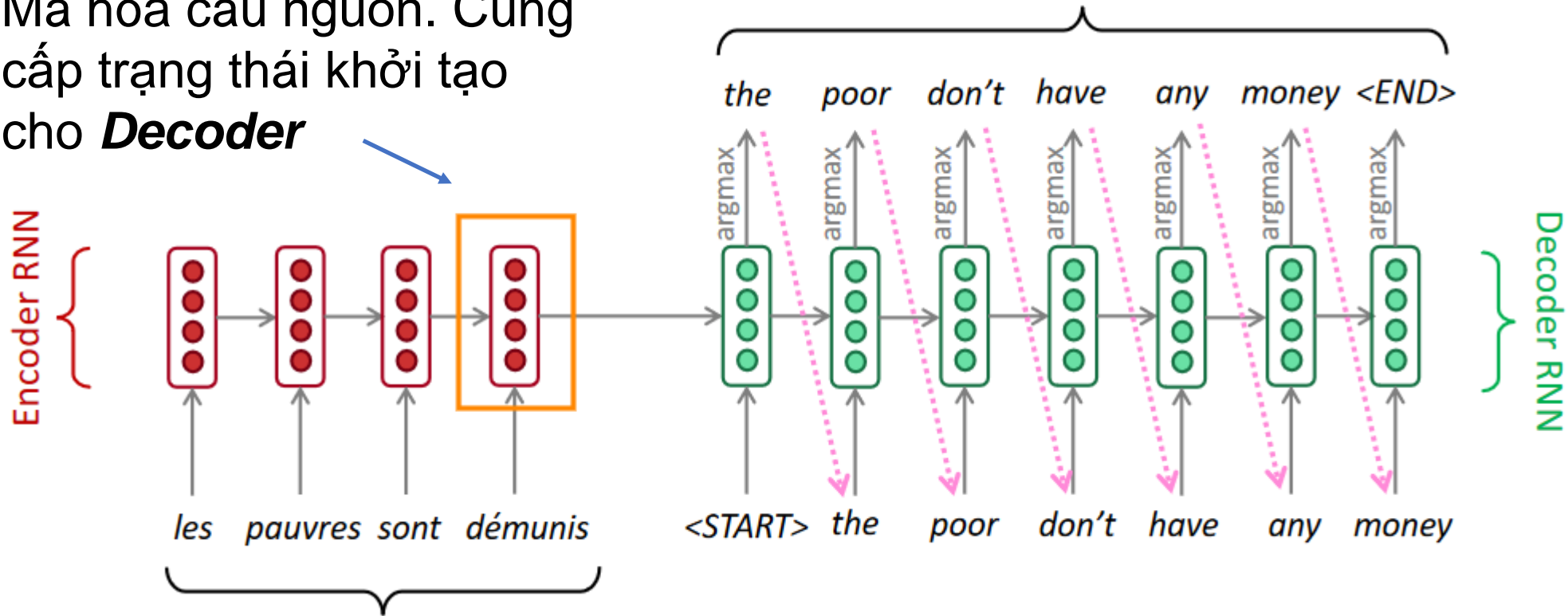
# Mô hình sequence to sequence

---

- Gồm 2 RNNs khác nhau để mô hình  $\prod_{i=1}^m p(e_i | e_1, \dots, e_{i-1}, f)$
- RNN đầu tiên gọi là encoder, mã hóa câu f (tiếng Pháp)
- RNN thứ 2 gọi là decoder, sinh câu e (Tiếng Anh)

# Mô hình Seq2Seq

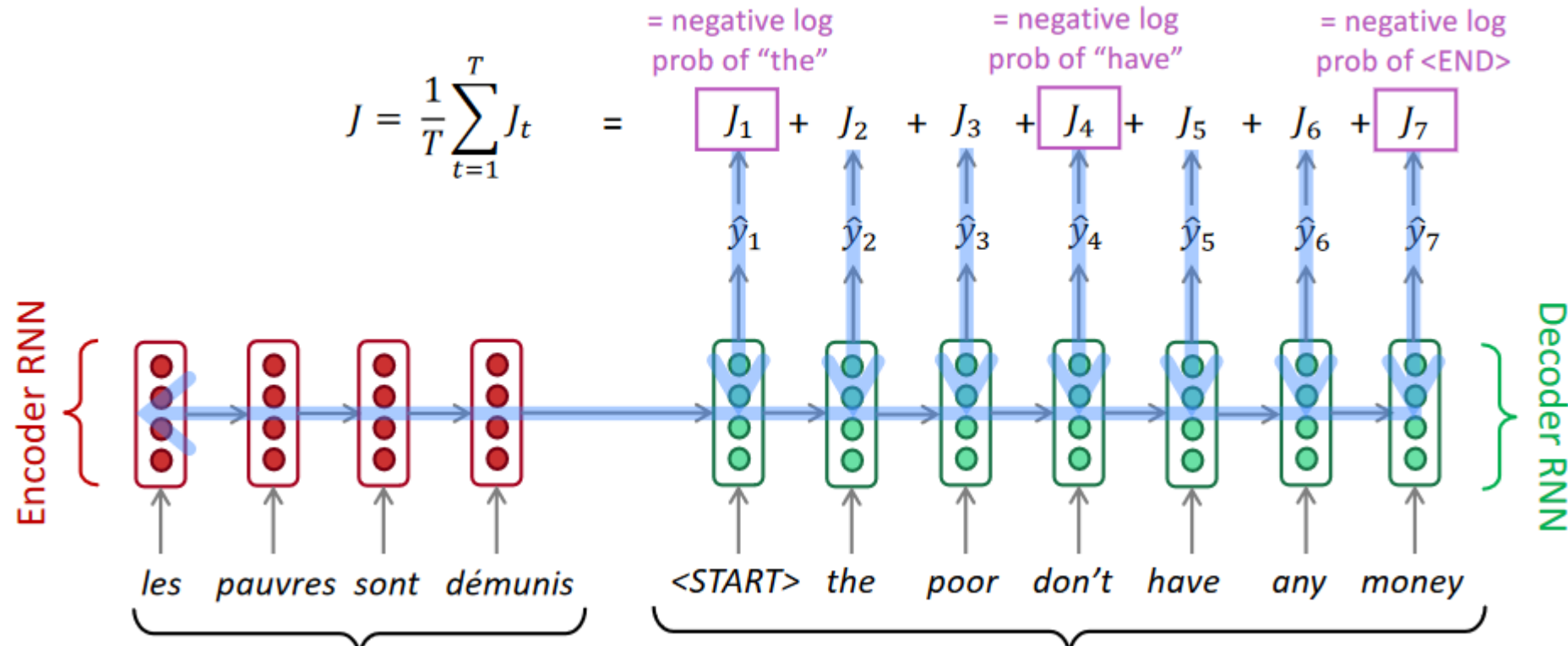
- Mã hóa câu nguồn. Cung cấp trạng thái khởi tạo cho **Decoder**



1) **Encoder**: Nén câu nguồn thành **vector**

2) **Decoder**: Là mô hình ngôn ngữ, sinh câu đích với điều kiện **vector nguồn**

# Huấn luyện mô hình Seq2Seq

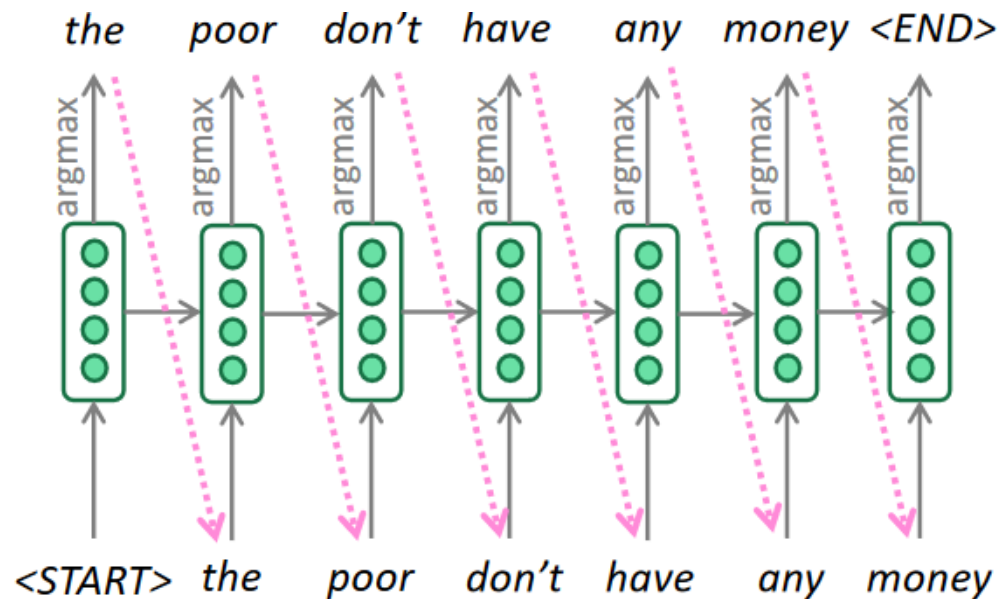


- Tham số của mô hình

$$W^{enc}, W^{dec}, W_{out}$$

# Bước giải mã

- Với mô hình Seq2Seq đã huấn luyện: Cho 1 câu tiếng Pháp, cần xác định câu tiếng Anh có xác suất cao nhất.
- Tìm  $\arg \max \prod_{i=1}^m p(e_i | e_1, \dots, e_{i-1}, f)$
- 
- Giải pháp đơn giản nhất: giải mã tham lam



Vấn đề gặp phải?



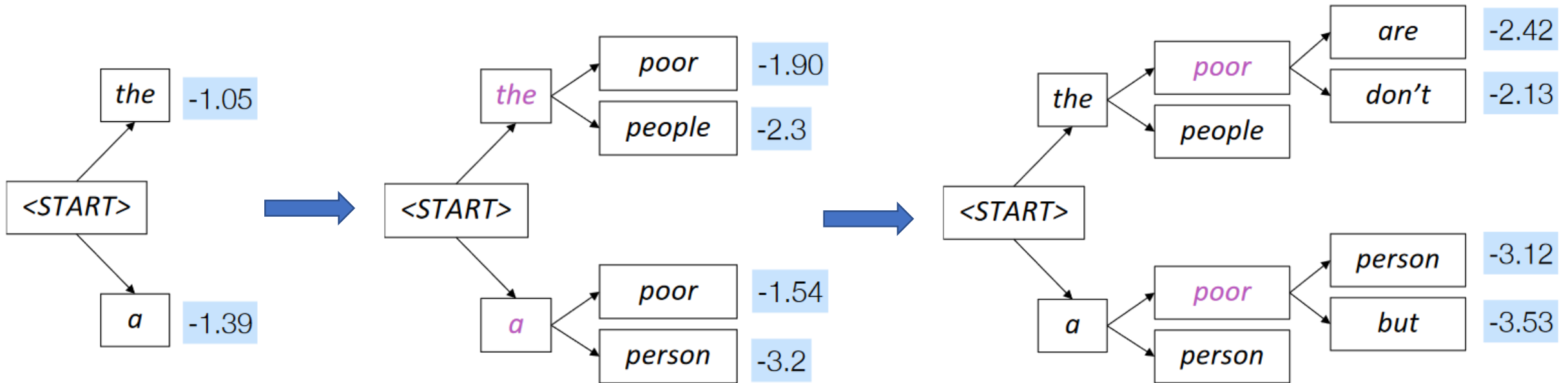
# Dùng beam search

---

- Giải mã tham lam: Chúng ta không thể quay lại để kiểm tra quyết định của bước trước.
  - les pauvres sont démunis (the poor don't have any money)
  - → the \_\_\_\_\_
  - → the poor \_\_\_\_\_
  - → the poor **are** \_\_\_\_\_
- Ý tưởng: Sử dụng beam search.
  - Lưu k bản dịch tại mỗi thời điểm thay vì chỉ 1
  - Thường beam size k khoảng 5 - 10

# Ví dụ beam search

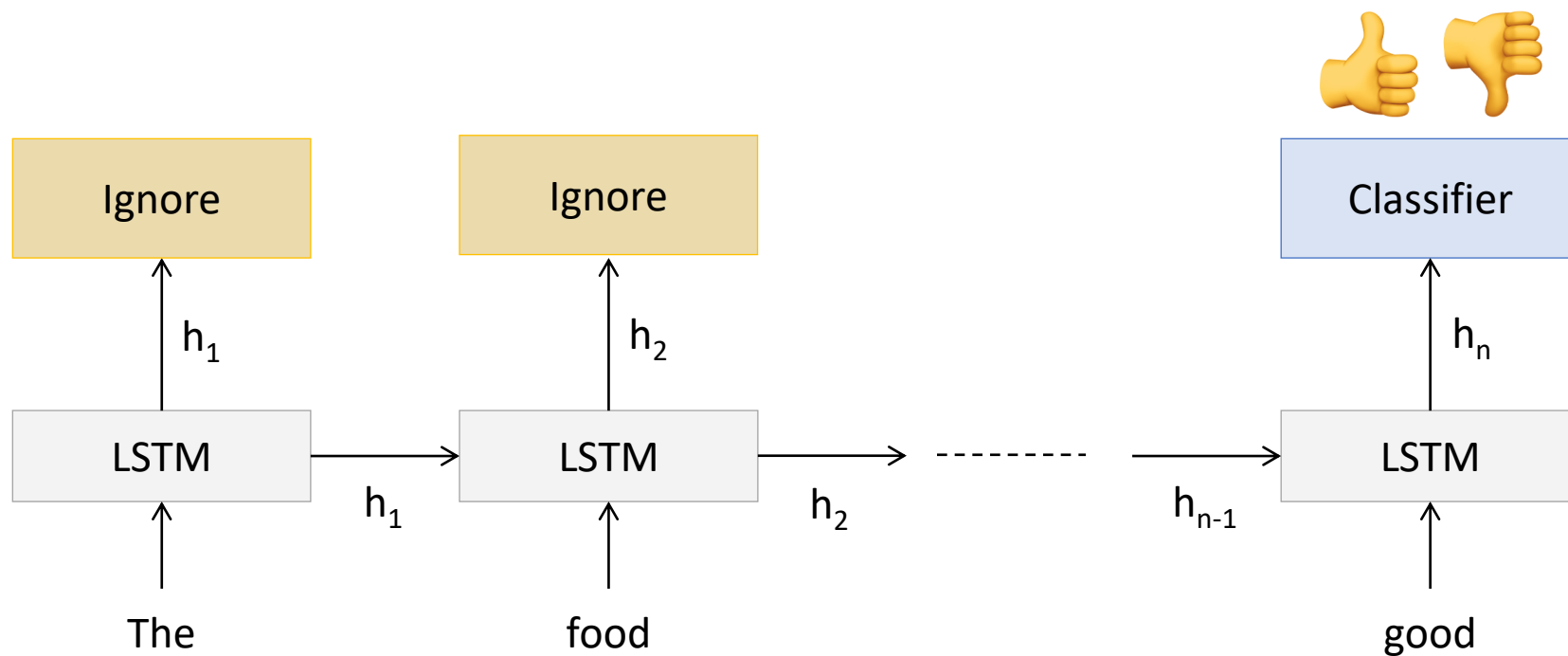
- Beam size = 2



# LSTM cho phân loại

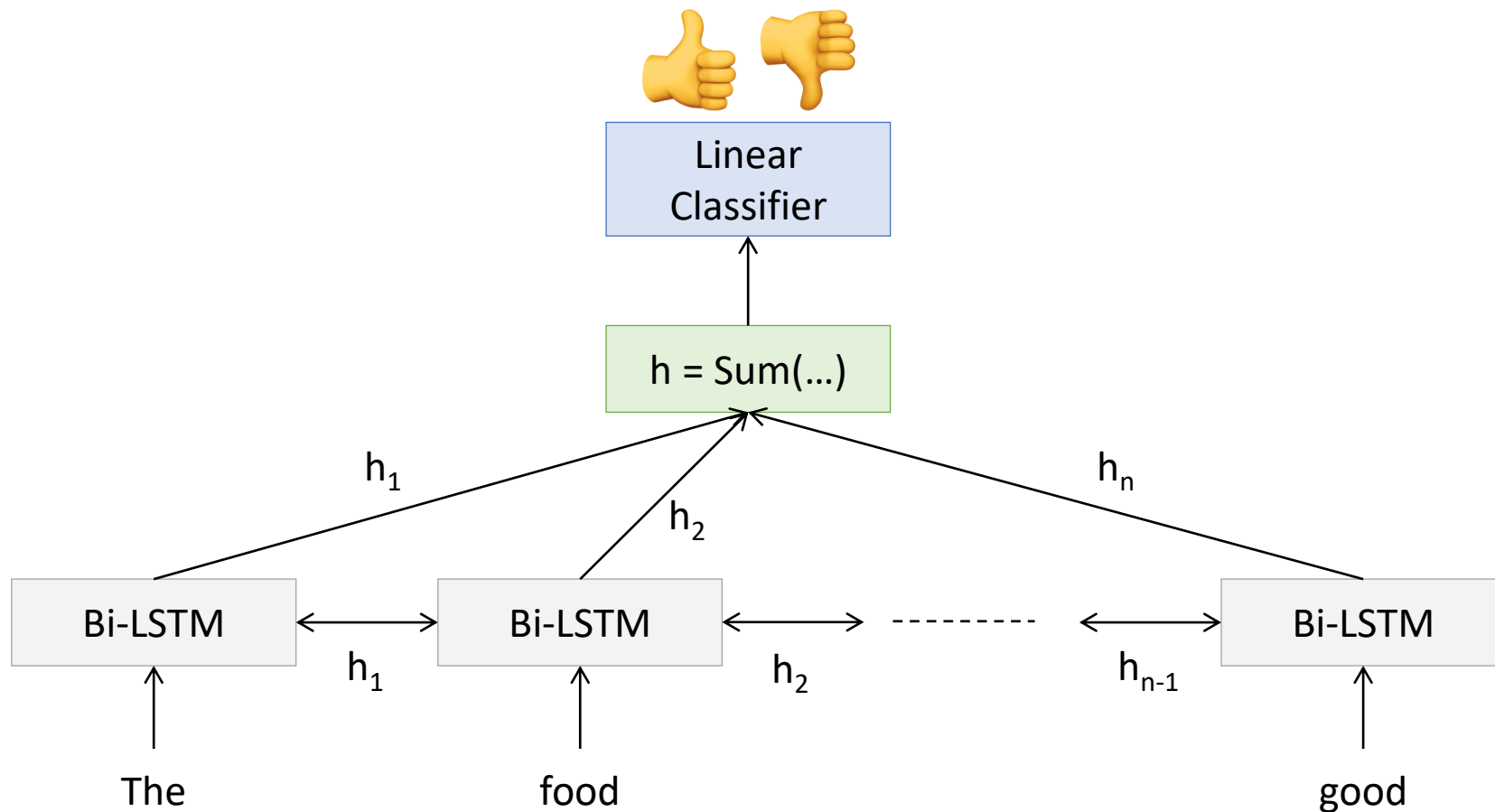
# Phân loại quan điểm với LSTM

- Đầu ra của từ cuối cùng được sử dụng làm đặc trưng để xây dựng bộ phân lớp



# Phân loại quan điểm với LSTM

- Đặc trưng được tính dựa trên đầu ra các từ



# LSTM cho gán nhãn chuỗi

# Một số bài toán

---

- Gán nhãn từ loại (Part Of Speech Tagging)
  - Mỗi từ trong câu được gán với 1 từ loại  
John saw the saw and decided to take it to the table.  
NNP VBD DT NN CC VBD TO VB PRP IN DT NN
  - Hữu ích trong phân tích cú pháp chuỗi con và phân biệt nghĩa của từ

# Một số bài toán

---

- Trích chọn thông tin

- Xác định cụm từ trong ngôn ngữ liên quan đến một nội dung cụ thể.
- Trích thông tin trong quảng cáo xe hơi

make model year mileage price

– For sale, 2002 Toyota Prius, 20,000 mi, \$15K or best offer.  
Available starting July 30, 2006.

- Xác định tên (Named entity recognition)

people organizations places

– Michael Dell is the CEO of Dell Computer Corporation and lives  
in Austin Texas.



# Một số bài toán

---

- Semantic Role Labeling

- Với mỗi vế, xác định vai trò ngữ nghĩa của mỗi cụm danh từ so với động từ

agent patient source destination instrument

– John drove Mary from Austin to Dallas in his Toyota Prius.

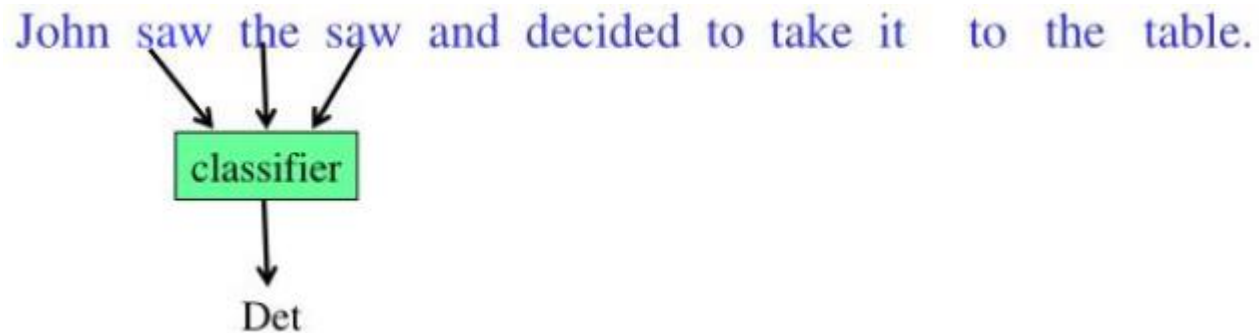
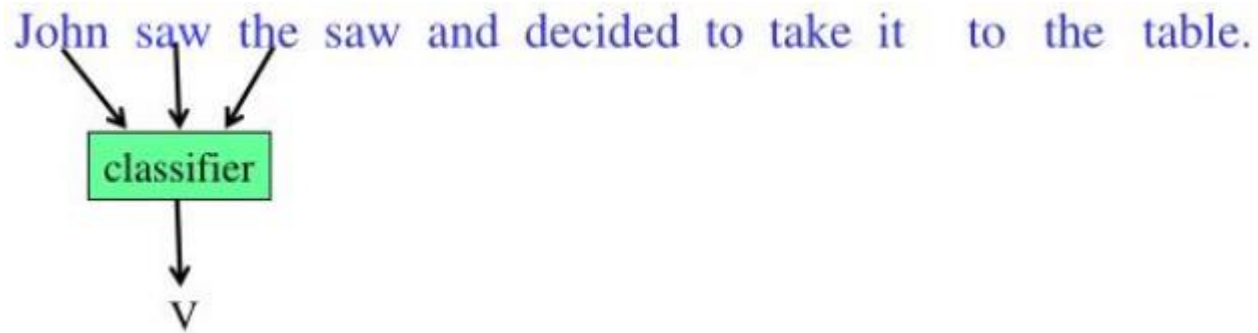
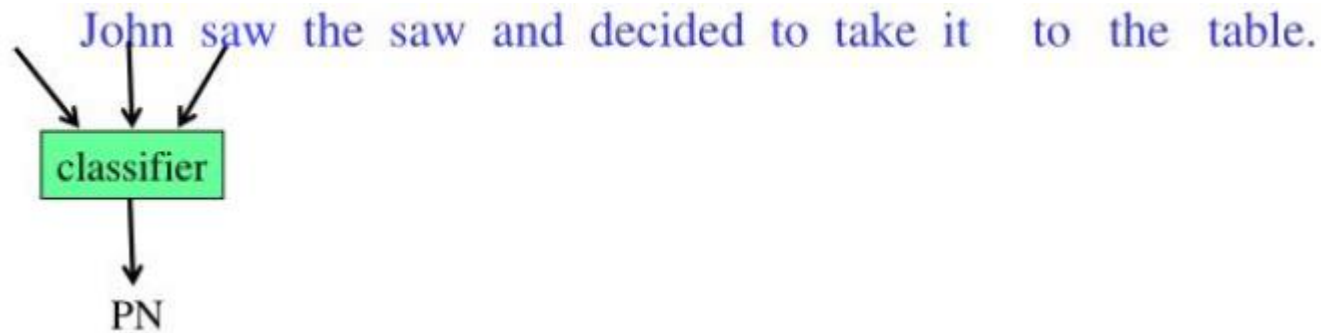
– The hammer broke the window.

- Gán nhãn cho chuỗi gen trong phân tích bộ gen

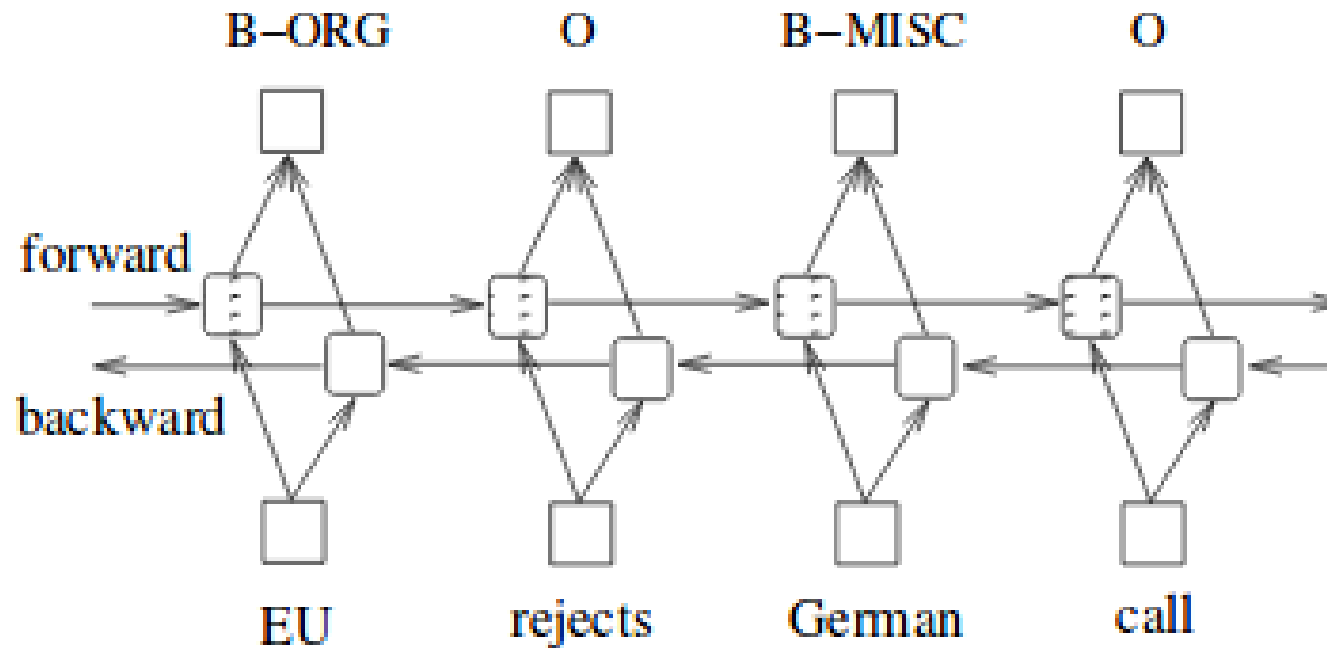
exon intron

– AGCTAACGTTTCGATACGGATTACAGCCT

# Gán nhãn chuỗi sử dụng phân lớp



# Sử dụng Bi-LSTM



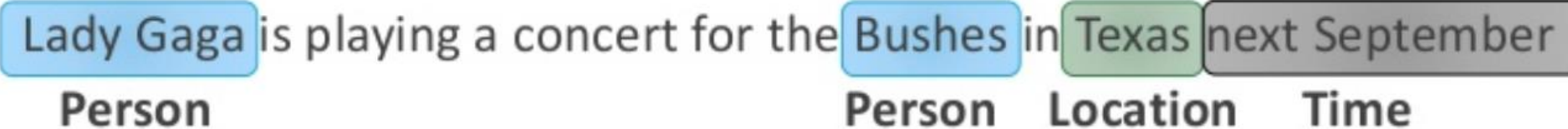
# Sử dụng LSTM cho NER

# Nhận dạng thực thể

---

- Mục đích: Xác định các tên trong văn bản và phân chúng vào các nhóm.
- Các nhóm: Người, vị trí, tổ chức
- Các bài toán khác: ngày, tháng, địa chỉ email, các độ đo
- Thực thể các lĩnh vực khác: tên thuốc, gen, mục tham chiếu

# Nhận dạng thực thể

- Ví dụ 
  - Person
  - Person
  - Location
  - Time
- Ứng dụng của NER:
  - Dịch máy
  - Hỏi – đáp
  - Truy vấn thông tin
  - Tổng hợp giọng nói

# Khó khăn

---

- Nhiều loại NEs: John Smith, Mr Smith, John.
- Nhập nhằng các kiểu NEs:
  - John Smith (công ty vs. người)
  - May (Người vs. tháng)
  - Washington (Người vs. địa điểm)
- Nhập nhằng với từ khác, ví dụ “may”
- Các kiểu, cấu trúc, dấu câu, định dạng

# Phương pháp học máy cho NER

---

- Huấn luyện
  - Thu thập dữ liệu
  - Gán nhãn mỗi từ là lớp thực thể của nó hoặc khác (O)
  - Trích chọn đặc trưng cho dữ liệu văn bản và các lớp
  - Huấn luyện bộ phân lớp

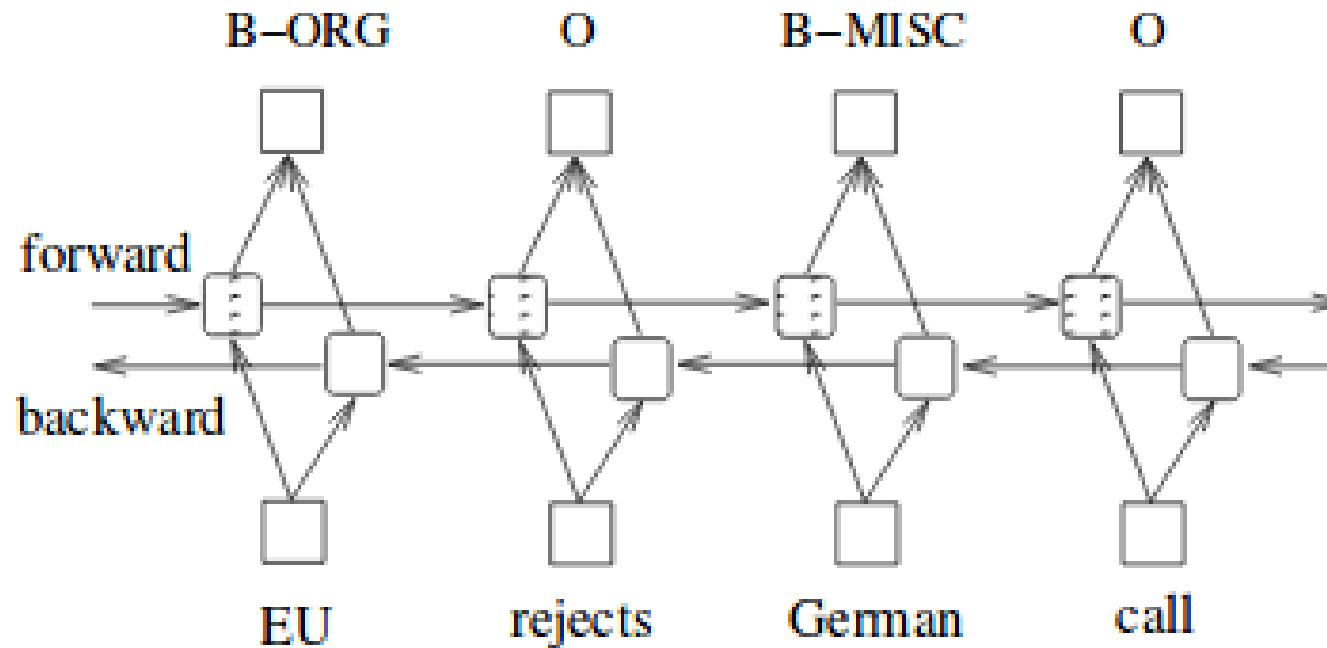


# Phương pháp học máy cho NER

- Gán nhãn lớp cho các từ

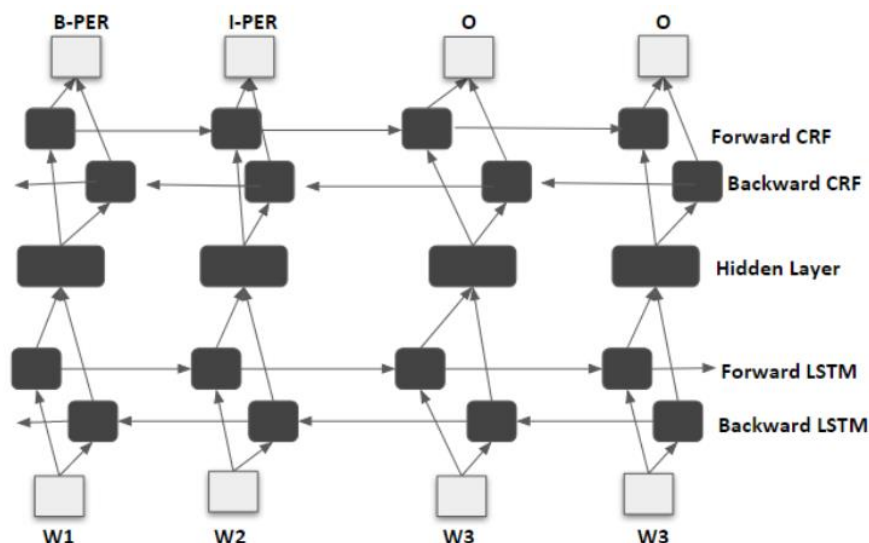
	IO encoding	IOB encoding
Fred	PER	B-PER
showed	O	O
Sue	PER	B-PER
Mengqiu	PER	B-PER
Huang	PER	I-PER
's	O	O
new	O	O
painting	O	O

# Sử dụng Bi-LSTM



# Mô hình BiLSTM-CRF cho bài toán NER/POS

- Gán nhãn chuỗi: BiLSTM được kết hợp với CRF để nâng cao kết quả
- So sánh mô hình sử dụng/không sử dụng CRF? Vai trò của CRF trong mô hình?



Mô hình LSTM-CRF cho gán nhãn chuỗi

System	accuracy
Maximum entropy cyclic dependency network (Toutanova et al., 2003)	97.24
SVM-based tagger (Gimenez and Marquez, 2004)	97.16
Bidirectional perceptron learning (Shen et al., 2007)	97.33
Semi-supervised condensed nearest neighbor (Soegaard, 2011)	97.50
CRFs with structure regularization (Sun, 2014)	97.36
Conv network tagger (Collobert et al., 2011)	96.37
Conv network tagger (senna) (Collobert et al., 2011)	97.29
BI-LSTM-CRF (ours)	<b>97.43</b>
BI-LSTM-CRF (Senna) (ours)	<b>97.55</b>

Kết quả so sánh cho bài toán POS trên tập Penn TreeBank (PTB)

# Q&A

Thank you!