Shape, arrow

Description automatically generatedShape, arrow

Description automatically generatedShape, arrow

Description automatically generated

BỘ GIÁO DỤC VÀ ĐÀO TẠO

TRƯỜNG ĐẠI HỌC SƯ PHẠM TP.HCM

Khoa Công nghệ Thông tin

-------🙣🙣🏵🙡🙡-------



BÁO CÁO ĐỀ TÀI

NGHIÊN CỨU KHOA HỌC

Giảng viên hướng dẫn : ThS. LƯƠNG TRẦN HY HIẾN

Nhóm thực hiện :

+ LÊ VĂN TRUNG 4501104259

+ NGUYỄN VĂN PHONG 4501104175

Shape, arrow

Description automatically generated

**TP HỒ CHÍ MINH – 2021**

Mục lục

[DANH MỤC HÌNH ẢNH 3](#_Toc68810685)

[LỜI CẢM ƠN 4](#_Toc68810686)

[CHƯƠNG 1: TỔNG QUAN 5](#_Toc68810687)

[1.1 Giới thiệu bài toán – Đặt vấn đề 5](#_Toc68810688)

[CHƯƠNG 2: cơ sở lý thuyết 6](#_Toc68810689)

[2.1 Dự báo chuỗi thời gian 6](#_Toc68810690)

[2.2 Deep-learning 6](#_Toc68810691)

[2.3 LSTM (Long short Time Memory) 7](#_Toc68810692)

[Chương 3: Xây dựng mô hình 8](#_Toc68810693)

[3.1 Thuật toán 8](#_Toc68810694)

[Tham số được dùng để dự đoán 11](#_Toc68810695)

# DANH MỤC HÌNH ẢNH

# LỜI CẢM ƠN

Nghiên cứu được thực hiện tại Khoa Công nghệ thông tin – Trường Đại học Sư phạm Thành phố Hồ Chí Minh, dưới sự hướng dẫn khoa học của ThS Lương Trần Hy Hiến.

Trước tiên chúng em xin bày tỏ lòng biết ơn sâu sắc tới thầy ThS Lương Trần Hy Hiến đã đưa chúng em đến với lĩnh vực nghiên cứu này. Thầy đã tận tình giảng dạy, hướng dẫn chúng em tiếp cận và đạt được những kết quả nhất định trong nghiên cứu của mình. Thầy đã luôn tận tâm động viên, khuyến khích và chỉ dẫn giúp chúng em hoàn thành nghiên cứu này.

Chúng em xin bày tỏ lòng biết ơn tới các Thầy Cô thuộc Khoa Công nghệ thông tin và cán bộ Phòng Đào tạo – Trường Đại học Sư Phạm Thành phố Hồ Chí Minh đã tạo mọi điều kiện thuận lợi giúp đỡ chúng em trong quá trình học tập và nghiên cứu. .

Lời đầu đầu tiên chúng em xin chân thành cảm ơn Khoa Công nghệ Thông tin – Trường Đại học Sư phạm Thành phố Hồ Chí Minh đã tạo điều kiện và giúp đỡ cho chúng em thực hiện đề tài nghiên cứu này.

Em xin gửi lời cảm ơn sâu sắc tới Thầy – Th.s Lương Trần Hy Hiến đã tận tình hướng dẫn, chỉ bảo cho chúng em trong suốt thời gian thực hiện đề tài.

Sự động viên, cổ vũ của gia đình, bạn bè là nguồn động lực quan trọng để chúng em thực hiện đề tài nghiên cứu.

Chúng em xin chân thành cảm ơn.

Do kiến thức còn hạn chế, nên đề tài nghiên cứu của chúng em không tránh khỏi những thiếu sót, kính mong sự thông cảm, chỉ bảo của quý Thầy Cô.

**Sinh viên thực hiện**

Lê Văn Trung

Nguyễn Văn Phong

# CHƯƠNG 1: TỔNG QUAN

## Giới thiệu bài toán – Đặt vấn đề

Dự báo chuỗi thời gian (time series forecasting) là một lĩnh vực quan trọng của Machine Learning thường bị bỏ qua. Dự báo chuỗi thời gian sử dụng các công nghệ khác nhau như Học máy, Mạng nơ ron nhân tạo, máy vectơ hỗ trợ, logic mờ, quy trình Gaussian và các mô hình Markov ẩn. Dữ liệu chuỗi thời gian có rất nhiều trong lĩnh vực thực tế. Việc nghiên cứu và phân tích chuỗi thời gian giúp đưa ra được các quyết định quan trọng. Nghiên cứu như vậy rất có hữu ích trong nhiều ứng dụng thực tế như thời tiết, doanh số, giá thành sản phẩm. Bài toán dự đoán giá vàng cũng là một dạng của dự báo chuỗi thời gian. Dự đoán giá vàng chính xác có thể giúp ích rất nhiều cho việc quản lí, phân phối vàng của chính phủ và các nhà đầu tư, liên hệ đến các thị trường khác như tiền tệ, giá thành sản phẩm. Cùng với sự phát triển của kỹ thuật Deep learning gần đây khiến cho việc giải quyết bài toán dự đoán giá vàng trở nên nhanh và chính xác hơn.

**Mục tiêu cụ thể:**

Xây dựng bộ dữ liệu tốt và đầy đủ nhằm hỗ trợ việc dự đoán trở nên chính xác hơn.

Dự đoán giá vàng của một ngày sau với bộ dữ liệu là giá vàng của thế giới từ năm 2014 - 2021.

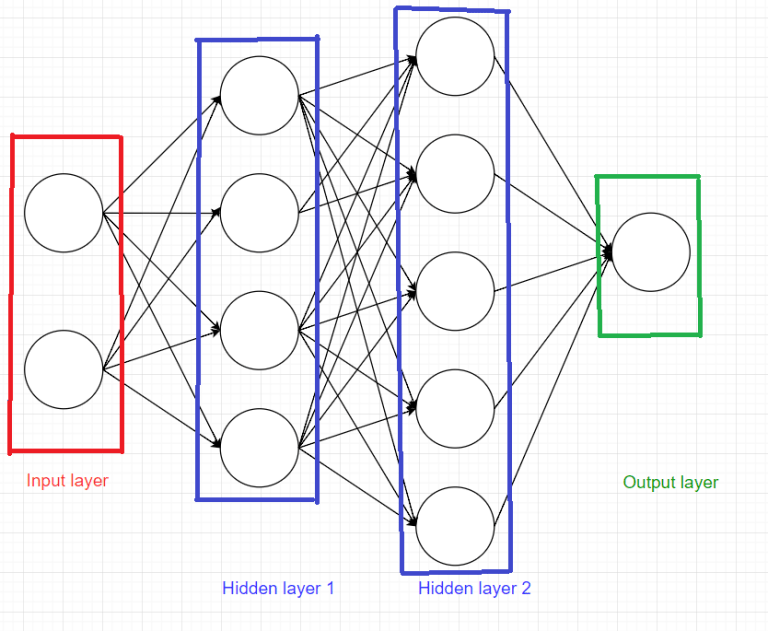
Mục tiêu cụ thể của đề tài là đưa ra được kết quả có độ chính xác 90% trở lên.

# CHƯƠNG 2: cơ sở lý thuyết

## 2.1 Dự báo chuỗi thời gian

Chuỗi thời gian (tiếng Anh: time series) trong thống kê, xử lý tín hiệu, kinh tế lượng và toán tài chính là một chuỗi các điểm dữ liệu, được đo theo từng khoảng khắc thời gian liền nhau theo một tần suất thời gian thống nhất. Ví dụ về chuỗi thời gian là giá đóng cửa của chỉ số Dow Jones hoặc lưu lượng hàng năm của sông Nin tại Aswan. Phân tích chuỗi thời gian bao gồm các phương pháp để phân tích dữ liệu chuỗi thời gian, để từ đó trích xuất ra được các thuộc tính thống kê có ý nghĩa và các đặc điểm của dữ liệu. Dự đoán chuỗi thời gian là việc sử dụng mô hình để dự đoán các sự kiện thời gian dựa vào các sự kiện đã biết trong quá khứ để từ đó dự đoán các điểm dữ liệu trước khi nó xảy ra (hoặc được đo). Chuỗi thời gian thường được vẽ theo các đồ thị.

## 2.2 Deep-learning

Học sâu hay Thâm học (tiếng Anh: deep learning) là một chi của ngành máy học dựa trên một tập hợp các thuật toán để cố gắng mô hình dữ liệu trừu tượng hóa ở mức cao bằng cách sử dụng nhiều lớp xử lý với cấu trúc phức tạp, hiện tại người ta thường sử dụng mạng Neron nhân tạo (Neural Network), hoặc bằng cách khác bao gồm nhiều hàm kích hoạt phi tuyến (vd: sigmoid, tanh…) Đây cũng là một trong những hứa hẹn của học sâu là thay thế các tính năng thủ công bằng các thuật toán hiệu quả đối với học không có giám sát hoặc nửa giám sát và tính năng phân cấp. Cụ thể là ta đặt một chuỗi thời gian có thể biểu diễn bằng nhiều vector của các giá trí đầu vào là các mô hình tiếp cận khác nhau (vd: RNN, CNN, LSTM) để đạt được kết quả mong muốn.

## 2.3 LSTM (Long short Time Memory)

# Chương 3: Xây dựng mô hình

## 3.1 Thuật toán

 Chìa khóa của LSTM là trạng thái tế bào (cell state) - chính đường chạy thông ngang phía trên của sơ đồ hình vẽ. Trạng thái tế bào là một dạng giống như băng truyền. Nó chạy xuyên suốt tất cả các mắt xích (các nút mạng) và chỉ tương tác tuyến tính đôi chút. Vì vậy mà các thông tin có thể dễ dàng truyền đi thông suốt mà không sợ bị thay đổi.

LSTM có khả năng bỏ đi hoặc thêm vào các thông tin cần thiết cho trạng thái tế báo, chúng được điều chỉnh cẩn thận bởi các nhóm được gọi là cổng (gate).

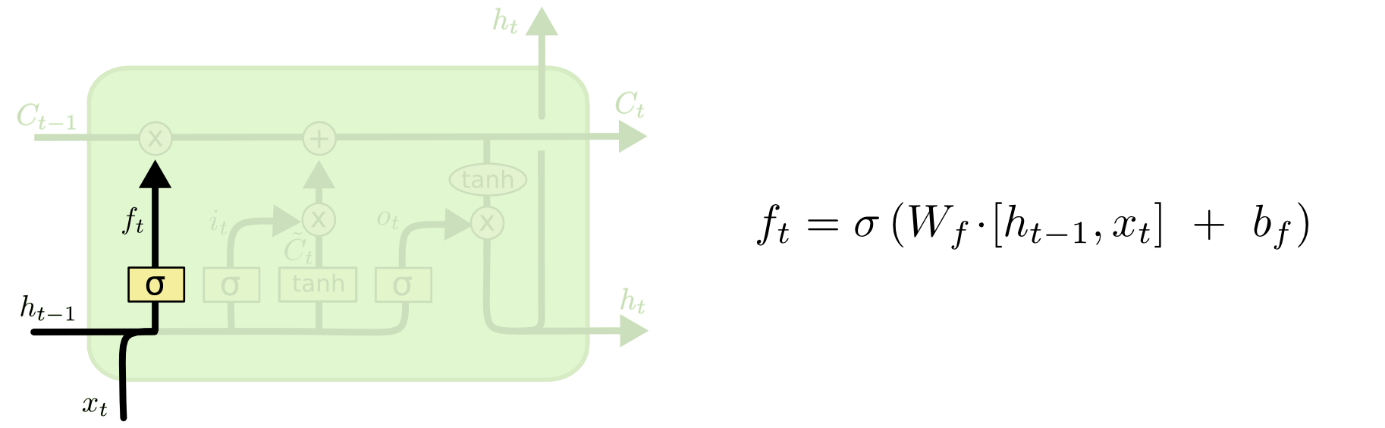
Các cổng là nơi sàng lọc thông tin đi qua nó, chúng được kết hợp bởi một tầng mạng sigmoid và một phép nhân.

Tầng sigmoid sẽ cho đầu ra là một số trong khoản [0, 1], mô tả có bao nhiêu thông tin có thể được thông qua. Khi đầu ra là 0 thì có nghĩa là không cho thông tin nào qua cả, còn khi là 1 thì có nghĩa là cho tất cả các thông tin đi qua nó.

Một LSTM gồm có 3 cổng như vậy để duy trì và điều hành trạng thái của tế bào.

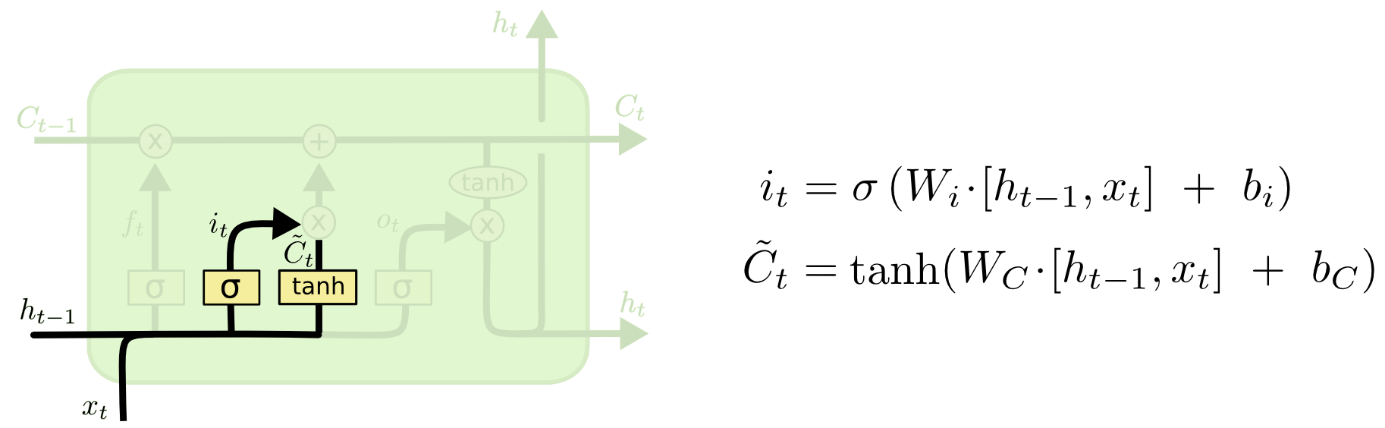
Bước đầu tiên của LSTM là quyết định xem thông tin nào cần bỏ đi từ trạng thái tế bào. Quyết định này được đưa ra bởi tầng sigmoid - gọi là “tầng cổng quên” (forget gate layer). Nó sẽ lấy đầu vào là  ​và  ​rồi đưa ra kết quả là một số trong khoảng [0, 1] cho mỗi số trong trạng thái tế bào  ​. Đẩu ra là 1 thể hiện rằng nó giữ toàn bộ thông tin lại, còn 0 chỉ rằng taonf bộ thông tin sẽ bị bỏ đi.

Quay trở lại với ví dụ mô hình ngôn ngữ dự đoán từ tiếp theo dựa trên tất cả các từ trước đó, với những bài toán như vậy, thì trạng thái tế bào có thể sẽ mang thông tin về giới tính của một nhân vật nào đó giúp ta sử dụng được đại từ nhân xưng chuẩn xác. Tuy nhiên, khi đề cập tới một người khác thì ta sẽ không muốn nhớ tới giới tính của nhân vật nữa, vì nó không còn tác dụng gì với chủ thế mới này.

[](https://colah.github.io/posts/2015-08-Understanding-LSTMs/img/LSTM3-focus-f.png)

Bước tiếp theo là quyết định xem thông tin mới nào ta sẽ lưu vào trạng thái tế bào. Việc này gồm 2 phần. Đầu tiên là sử dụng một tầng sigmoid được gọi là “tầng cổng vào” (input gate layer) để quyết định giá trị nào ta sẽ cập nhập. Tiếp theo là một tầng  tạo ra một vector cho giá trị mớinhằm thêm vào cho trạng thái. Trong bước tiếp theo, ta sẽ kết hợp 2 giá trị đó lại để tạo ra một cập nhập cho trạng thái.

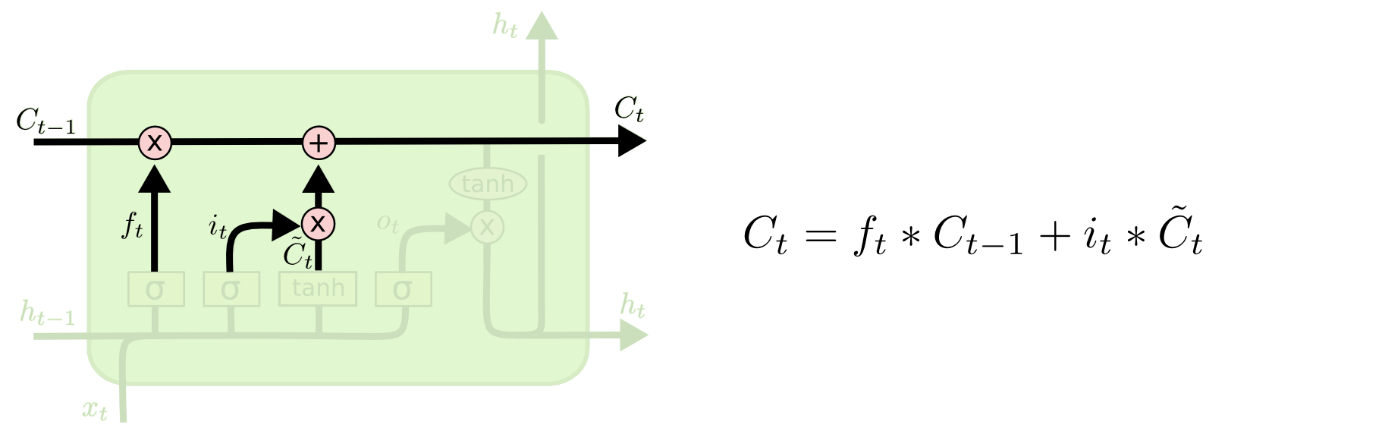
Chẳng hạn với ví dụ mô hình ngôn ngữ của ta, ta sẽ muốn thêm giới tính của nhân vật mới này vào trạng thái tế bào và thay thế giới tính của nhân vật trước đó.

[](https://colah.github.io/posts/2015-08-Understanding-LSTMs/img/LSTM3-focus-i.png)

Giờ là lúc cập nhập trạng thái tế bào cũ ​ thành trạng thái .Ở các bước trước đó đã quyết định những việc cần làm, nên giờ ta chỉ cần thực hiện là xong.

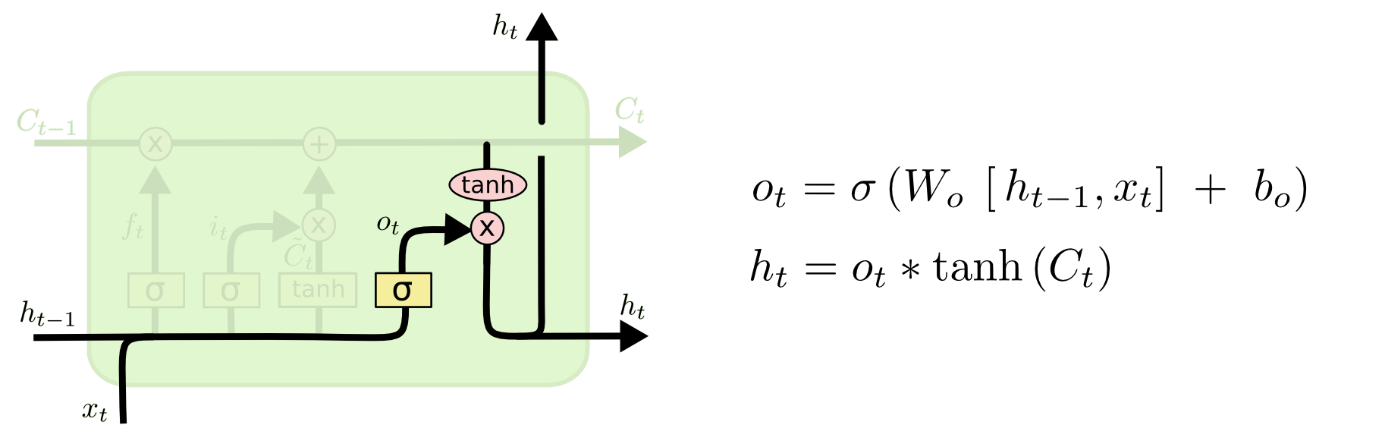
Ta sẽ nhân trạng thái cũ với ​ để bỏ đi những thông tin ta quyết định quên lúc trước. Sau đó cộng thêm  \* ​. Trạng thái mơi thu được này phụ thuộc vào việc ta quyết định cập nhập mỗi giá trị trạng thái ra sao.

Với bài toàn mô hình ngôn ngữ, chính là việc ta bỏ đi thông tin về giới tính của nhân vật cũ, và thêm thông tin về giới tính của nhân vật mới như ta đã quyết định ở các bước trước đó.

[](https://colah.github.io/posts/2015-08-Understanding-LSTMs/img/LSTM3-focus-C.png)

Cuối cùng, ta cần quyết định xem ta muốn đầu ra là gì. Giá trị đầu ra sẽ dựa vào trạng thái tế bào, nhưng sẽ được tiếp tục sàng lọc. Đầu tiên, ta chạy một tầng sigmoid để quyết định phần nào của trạng thái tế bào ta muốn xuất ra. Sau đó, ta đưa nó trạng thái tế bảo qua một hàm tanh để co giá trị nó về khoảng [-1, 1], và nhân nó với đầu ra của cổng sigmoid để được giá trị đầu ra ta mong muốn.

Với ví dụ về mô hình ngôn ngữ, chỉ cần xem chủ thể mà ta có thể đưa ra thông tin về một trạng từ đi sau đó. Ví dụ, nếu đầu ra của chủ thể là số ít hoặc số nhiều thì ta có thể biết được dạng của trạng từ đi theo sau nó phải như thế nào.

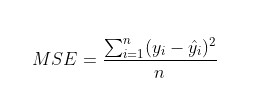
[](https://colah.github.io/posts/2015-08-Understanding-LSTMs/img/LSTM3-focus-o.png)

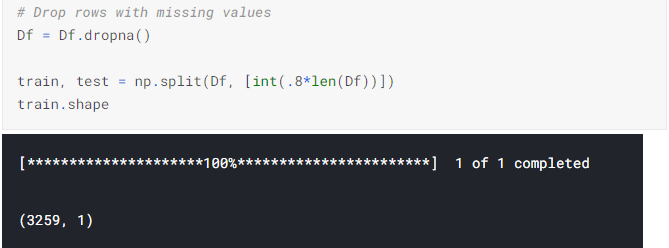
## Tham số được dùng để dự đoán

-hàm mean square error:

Ưu điểm: Xử lý tốt đạo hàm

Nhược điểm:Các dự đoán xa giá trị thực tế sẽ chênh lẹch lớn so với các giá trị gần đúng

1. Tập train, test



Đây là 2 tham số để chia mô hình làm 2 phần: train, test.Trong đó tập train