

**TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM
TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG
KHOA CÔNG NGHỆ THÔNG TIN**

_____ * _____



BÁO CÁO MÔN KIẾN TẬP CÔNG NGHIỆP

**HỌC BIỂU DIỄN TƯỢNG PHẢN
GIAI ĐOẠN 1: KHẢO SÁT CÁC MÔ HÌNH HỌC
BIỂU DIỄN TƯỢNG PHẢN**

Cán bộ hướng dẫn: **PGS. TS. Nguyễn Thanh Hiên**

Giảng viên giám sát: **TS. Huỳnh Ngọc Tú**

Người thực hiện: **Trần Quốc Lĩnh - 51703124**

TP. Hồ Chí Minh, ngày 20 tháng 08 năm 2020

LỜI CẢM ƠN

Em xin chân thành cảm ơn khoa Công Nghệ Thông Tin và trường Đại Học Tôn Đức Thắng và Công ty TNHH Tin học Đại Phát. Với sự giúp đỡ của trường, khoa và công ty, em đã hoàn thành báo cáo môn dự án công nghệ thông tin 1 với đề tài: Tìm hiểu về time series forecasting trong deep learning (phần 2)

Trong quá trình soạn thảo không thể tránh khỏi những thiếu sót. Em rất mong nhận được ý kiến đóng góp của các thầy cô để hoàn thiện tốt hơn báo cáo của em. Và cũng như để nâng cao kiến thức, kinh nghiệm của bản thân. Xin chân thành cảm ơn!

TP. Hồ Chí Minh, Ngày 20 tháng 08 năm 2020

Tác giả

BÀI BÁO CÁO ĐƯỢC HOÀN THÀNH TẠI TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG

Em xin cam đoan đây là sản phẩm nghiên cứu của riêng em. Các nội dung nghiên cứu, kết quả trong đề tài này là trung thực và chưa công bố dưới bất kỳ hình thức nào trước đây. Những số liệu, hình ảnh được chính em thu thập từ các nguồn khác nhau có ghi rõ trong phần tài liệu tham khảo.

Ngoài ra, trong bài tiểu luận còn sử dụng một số nhận xét, đánh giá cũng như số liệu của các tác giả khác, cơ quan tổ chức khác đều có trích dẫn và chú thích nguồn gốc.

Nếu phát hiện có bất kỳ sự gian lận nào em xin hoàn toàn chịu trách nhiệm về nội dung bài tập lớn của mình. Trường đại học Tôn Đức Thắng không liên quan đến những vi phạm tác quyền, bản quyền do em gây ra trong quá trình thực hiện (nếu có).

TP. Hồ Chí Minh, Ngày 20 tháng 08 năm 2020

Tác giả

Trần Quốc Lĩnh

PHẦN XÁC NHẬN VÀ ĐÁNH GIÁ CỦA GIẢNG VIÊN

Phần xác nhận của GV hướng dẫn

.....

.....

.....

.....

.....

.....

TP. Hồ Chí Minh, ngày tháng năm
(kí và ghi họ tên)

Phần đánh giá của GV chấm bài

.....

.....

.....

.....

.....

.....

TP. Hồ Chí Minh, ngày tháng năm
(kí và ghi họ tên)

Tóm tắt

Ở bài báo cáo trước em đã trình bày một cái tổng quát và khái quát các nội dung về Time series forecasting. Thế nên, trong bài báo cáo này em sẽ trình bày một chút cụ thể hơn, chi tiết hơn thông qua các mô hình dự báo được ưa chuộng và hiệu quả nhất ở thời điểm hiện tại. Sau đó, đánh giá các mô hình trong một ví dụ thực tiễn.

Mục lục

1	Các mô hình time series forecasting	2
1.1	Autoregressive Integrated Moving Average (ARIMA)	2
1.2	Generalize Autoregressive Conditionally Hestoroscedastic (GARCH)	3
1.3	Long Short Term Memories (LSTM)	4
2	Đánh giá các mô hình dự báo	5
3	Tổng kết	6
	TÀI LIỆU THAM KHẢO	7

Danh sách hình vẽ

Chương 1

Các mô hình time series forecasting

Để thực hiện, xây dựng các mô hình chuỗi thời gian, hiện tại có hai ngôn ngữ hỗ trợ tốt cho việc này là R (R có các packages như forecast và lmtest) và python. Mặc dù python dễ sử dụng và có cộng đồng lớn mạnh hơn nhiều so với R, nhưng R lại hỗ trợ tốt hơn python trong thống kê và hiện thực hóa mô hình chuỗi thời gian. Đó cũng là một trong những lý do mà các nhà thống kê và kinh tế lượng ưa chuộng sử dụng R. Tuy nhiên, ở bài báo cáo này chúng ta chỉ dừng lại với python trong việc tìm hiểu về cách xây dựng các mô hình.

1.1 Autoregressive Integrated Moving Average (ARIMA)

Thông qua bài báo cáo trước, ắt hẳn chúng ta đã biết rằng chuỗi thời gian là các thông số được ghi lại theo thời gian, theo các mốc thời gian xác định. Và giá trị của hiện tại có sự tương quan đến các giá trị trong quá khứ, dễ hiểu hơn là để xác định giá trị trong thời điểm hiện tại cần có các giá trị trong quá khứ.

Mô hình ARIMA là một mô hình dựa trên ý tưởng dự đoán giá trị tương lai bằng các thông số đã ghi nhận được trong quá khứ. Và giả thuyết chuỗi thời gian là chuỗi dừng và phương sai của sai số không đổi

ARIMA gồm AR (Auto regression), I (Intergrated) kết hợp với MA (Moving Average). Trong đó:

- **Auto regression:** Có nghĩa là tự hồi qui. Thành phần hồi qui này gồm một tập hợp các giá trị lùi về p bước thời gian của chuỗi. Được biểu diễn dưới dạng:

$$AR(p) = \sum_{i=0}^p \phi x_{t-i} = \phi_0 + \phi x_{t-1} + \phi x_{t-2} + \dots + \phi x_{t-p}$$

- **Moving average:** Có nghĩa là trung bình trượt. Đây là quá trình dịch chuyển - quá trình thay đổi giá trị trung bình của chuỗi. Tuy nhiên chuỗi này phải thỏa mãn các tính chất sau:

$$\begin{cases} E(\varepsilon_t) = 0 & (0) \\ \sigma(\varepsilon_t) = \alpha & (2) \\ \rho(\varepsilon_t, \varepsilon_{t-s}) = 0, \forall s < t & (3) \end{cases}$$

Trong đó: (1) nghĩa là kỳ vọng phải bằng không và (2) phương sai không đổi để đảm bảo tính dừng của chuỗi.

Quá trình Moving average sẽ tìm mối liên hệ về mặt tuyến tính giữa các phần tử ngẫu nhiên ε_t của giá trị hiện tại và quá khứ (stochastic term). Có nhiều phương pháp để tính giá trị trung bình của chuỗi như WMA, EMA hay đơn giản nhất là Simple Moving Average (SMA):

$$MA(q) = \frac{\sum_{i=0}^q Value_{t-i}}{q}$$

- **Intergrated:** Là quá trình đồng tích hợp hoặc lấy sai phân. Hầu hết các chuỗi thời gian không có tính dừng thường. Do đó ta cần biến đổi nó sang chuỗi dừng bằng sai phân. Khi biến đổi sang chuỗi dừng, các nhân tố ảnh hưởng thời gian được loại bỏ và chuỗi sẽ dễ dự báo hơn. Quá trình sai phân bậc d của chuỗi được thực hiện như sau:

$$\begin{aligned} I(1) &= \Delta(x_t) = x_t - x_{t-1} \text{ (sai phân bậc 1)} \\ I(d) &= \Delta^d(x_t) = \underbrace{\Delta(\Delta(\dots \Delta(x_t)))}_{d \text{ times}} \text{ (sai phân bậc d)} \end{aligned}$$

Như vậy, tham số đặc trưng của mô hình được đặc tả bởi 3 tham số ARIMA(p, d, q). Và có thể được biểu diễn dưới dạng:

$$\Delta x_t = \phi_1 \Delta x_{t-1} + \phi_2 \Delta x_{t-2} + \dots + \phi_p \Delta x_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

1.2 Generalize Autoregressive Conditionally Heteroscedastic (GARCH)

GARCH (Generalize Autoregressive Conditionally Heteroscedastic) là một dạng mô hình thuộc nhóm hồi qui chuỗi thời gian thường được áp dụng phổ biến trong các dự báo kinh tế và tài chính. Ban đầu, mô hình gốc ARCH được giới thiệu bởi Engle (1982) để hồi qui các chuỗi dừng nhưng có mối quan hệ phi tuyến tính. Nguyên nhân của sự phi tuyến tính này xuất phát từ việc phương sai sai số của mô hình

thay đổi dẫn đến việc áp dụng các mô hình dựa trên giả định phương sai sai số không đổi như ARIMA hoặc ARMA không đạt kết quả chuẩn xác. Trong khi đó hiện tượng phương sai sai số thay đổi khá phổ biến trong các chuỗi chứng khoán và tài chính bởi luôn có những cú sốc kinh tế thường đến bất ngờ và là một nhân tố mà mô hình không thể giải thích được. Trong biểu diễn phân phối của chuỗi chúng ta sẽ thấy chuỗi có dạng đuôi béo (fat-tail). Điều đó cho thấy chuỗi có nhiều outlier tập trung ở các khoảng cận trên hoặc cận dưới. Bên cạnh đó một mô hình ARCH là một mô hình được xây dựng với biến mục tiêu là phương sai của một chuỗi số. Do đó ARCH thường được sử dụng để giải thích sự thay đổi hoặc sự biến động của phương sai chuỗi. Mặc dù ARCH có thể được sử dụng để mô tả sự tăng tiến của phương sai qua thời gian nhưng thay vào đó nó thường được sử dụng để mô tả sự gia tăng biến động chuỗi trong một khoảng thời gian ngắn. Trong kinh tế và tài chính mô hình ARCH thường được áp dụng với những chuỗi tăng dần (hoặc giảm dần) theo chu kì như giá chứng khoán, GDP, qui mô dân số, . . . Do đó chúng ta cần các biến để diễn tả được xu thế tăng (giảm) này nhưng vẫn đảm bảo được chuỗi dừng như là điều kiện cần của chuỗi thời gian. Vì thế các biến thường được sử dụng trong mô hình GARCH thường là: Phần trăm lãi / lỗ qua thời gian: Ngoài ra mô hình GARCH có thể sử dụng bất kì biến nào khác có yếu tố tăng giảm phương sai theo chu kì. Chẳng hạn như số dư từ chính mô hình ARIMA đã được hồi qui.

1.3 Long Short Term Memories (LSTM)

Chương 2

Đánh giá các mô hình dự báo

Chương 3

Tổng kết

Kết quả

Kết quả đạt được

Qua những nội dung được trình bày ở trên, ta có thể thấy bài báo cáo đã trình bày một cách cơ bản và đầy đủ nhất chủ đề tìm hiểu về time series forecasting. Qua đó ta có thể hiểu được thế nào là time series, ứng dụng và cách triển khai thuật toán. Nó mang nhiều giá trị hiện thực không chỉ trong nghiên cứu khoa học mà còn có thể áp dụng vào thực tiễn của cuộc sống.

Hạn chế

Bài báo cáo tuy đã trình bày tốt nhất có thể nhưng vẫn còn tồn tại một vài thiếu sót nhỏ. Các phần cần nêu theo mục đích của đề tài đã được hoàn thành nhưng chỉ ở mức là tổng quát, chung nhất, chưa đi vào chi tiết và trình bày cụ thể.

Hướng phát triển

Có nhiều hướng phát triển khác nhau cho các bài time series forecasting. Ngoài việc xây dựng thuật toán theo các phương pháp trên thì còn có nhiều phương pháp khác, được thiết kế lại để tối ưu hóa hiệu năng, độ chính xác,... cho những trường hợp và mục đích riêng biệt. Ngành công nghệ thông tin ngày càng phát triển, đặc biệt là các lĩnh vực trí tuệ nhân tạo, học máy, học sâu. Nên trong tương lai, chắc chắn sẽ có thêm nhiều cách tiếp cận mới, hiệu quả và chính xác hơn. Không chỉ hữu ích trong bài toán time series forecasting mà còn nhiều bài toán khác nữa.

Tài liệu tham khảo

- [1] ADAM HAYES. Simple moving average (sma).
- [2] Phạm Đình Khanh. Mô hình arima trong time series.