

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



BÁO CÁO KẾT THÚC MÔN: KIẾN TẬP CÔNG NGHIỆP

TÌM HIỂU VỀ TIME SERIES FORECASTING TRONG DEEP LEARNING (PHẦN 2)

Người hướng dẫn: Nguyễn Thanh Hiên

Người thực hiện: Trần Quốc Lĩnh - 51703124

Lớp: 17050301

Khoá: 21

THÀNH PHỐ HỒ CHÍ MINH, NĂM 2020

LỜI CẢM ƠN

Cảm ơn thầy Nguyễn Thanh Hiên đã hướng dẫn và tận tình chỉ bảo em trong suốt thời gian em thực tập ở công ty TNHH Tin học Đại Phát, cũng như trong suốt thời gian em thực hiện bài báo cáo này.

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

(Đã ký)

Trần Quốc Lĩnh

PHẦN XÁC NHẬN VÀ ĐÁNH GIÁ CỦA GIẢNG VIÊN

Phần xác nhận của GV hướng dẫn

Tp. Hồ Chí Minh, ngày 20 tháng 08 năm 2020
(*kí và ghi họ tên*)

Phần đánh giá của GV giám sát

Tp. Hồ Chí Minh, ngày 20 tháng 08 năm 2020
(*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

LỜI CẢM ƠN	2
CAM KẾT	3
ĐÁNH GIÁ CỦA GIÁO VIÊN	4
TÓM TẮT	5
MỤC LỤC	6
DANH MỤC CHÚ THÍCH CÁC THUẬT NGỮ VÀ HÌNH ẢNH	7
CHƯƠNG I: BỐI CẢNH HIỆN TẠI VÀ TỔNG QUAN VỀ TIME SERIES FORECASTING	8
Time series	9
Time series forecasting	10
Những phương pháp được sử dụng trong time series forecasting	10
CHƯƠNG II: MỘT SỐ ỨNG DỤNG THỰC TẾ	13
Dự đoán mức tăng trưởng của một loại cổ phiếu	13
Dự báo thời tiết	14
Dự đoán doanh thu	15
CHƯƠNG III: VÍ DỤ MINH HỌA DỰ BÁO THỜI TIẾT	17
CHƯƠNG IV: TỔNG KẾT	32
TÀI LIỆU THAM KHẢO	33

DANH MỤC CHÚ THÍCH

Hình ảnh

Hình 1:

Hình 2:

Hình 3:

Hình 4:

Hình 5:

Chương 1

CÁC MÔ HÌNH TIME SERIES FORECASTING

1.1 Autoregressive Integrated Moving Average (ARIMA)

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 lmttest) 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 hơn. Tuy nhiên, ở bài báo cáo này chúng ta chỉ dừng lại với python trong việc tìm hiểu về cách xây dựng mô hình ARIMA.

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 kỳ 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 kỳ. 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]. <https://towardsdatascience.com/deep-learning-for-time-series-and-why-deep>
- [2]. <https://www.slideshare.net/BeriDang/m-hnh-arima-103586847>
- [3]. <https://viblo.asia/p/time-series-data-gDVK2Qbv5Lj>
- [4]. https://en.wikipedia.org/wiki/Time_series
- [5]. <https://arxiv.org/pdf/2004.13408.pdf>
- [6]. Deep Learning for Time Series Forecasting - Jason Brownlee
- [7]. <http://www.ijmlc.org/vol7/632-P17.pdf>
- [8]. https://www.tensorflow.org/tutorials/structured_data/time_series
- [9]. <https://www.kaggle.com/phamdinhkhanh/garch-timeseries-model?fbclid=IwAR1fZCpLK8zbhS7guznWZmXgIjSEL-kqDfKdawXghcDXHHPd3w-sAXT6iJU>
- [10].
- [11].