

Meta-learning for multi foreign exchange movement prediction

Bao-Long Nguyen,^{1,*} Tom Ichibha,^{1,†} and Ryo Maezono^{1,‡}

¹*School of Information Science, JAIST, 1-1 Asahidai, Nomi, Ishikawa, 923-1292, Japan*

(Dated: August 22, 2024)

Abstract. Xử lý bài toán dự đoán các chỉ số liên quan đến tài chính bằng machine learning gặp rất nhiều khó khăn vì loại dữ liệu này không có chu kỳ rõ ràng và không chỉ phụ thuộc vào lịch sử giao dịch, mà còn phụ thuộc rất nhiều vào tình hình kinh tế, chính trị, điều mà rất khó capture trong tập dữ liệu. Để vượt qua các thách thức này, chúng tôi sử dụng Meta-learning, kết hợp với LSTM và CNN để tổng hợp hiệu quả các đặc trưng ẩn của dữ liệu theo thời gian. Thử nghiệm trên bài toán dự đoán xu hướng tỷ giá ngoại hối của 15 cặp tiền tệ cho thấy phương pháp của chúng tôi đạt hội tụ ở mức cao khi so sánh với mô hình NHITS.

I. INTRODUCTION

Foreign exchange prediction nói riêng và dự đoán các chỉ số tài chính như giá chứng khoán nói chung từ lâu đã là vấn đề đáng quan tâm của nhiều nghiên cứu [10, 13, 15]. Hai kỹ thuật chính được sử dụng trong financial time-series prediction là fundamental analysis and technical analysis [2]. Trong khi fundamental analysis thiên về phân tích các yếu tố như chính sách, chiến lược kinh tế của công ty, quốc gia để dự đoán tương lai; technical analysis dựa hoàn toàn vào lịch sử biến động giá để phân tích xu hướng thị trường.

Việc dự đoán các chỉ số tài chính này gặp một vài thách thức cố hữu. Trong đó có thể kể tới: (1) - Variance của loại dữ liệu này không hề cố định, từ đó giả thuyết rằng chúng tuân theo một phân phối để có thể xấp xỉ lỗi là không thể sử dụng được, dẫn đến việc các mô hình học máy khó có thể dự đoán chính xác giá cũng như xu hướng tăng giảm trong tương lai; (2) - Các chỉ số tài chính phần lớn không tuân theo bất kỳ quy luật rõ ràng nào nên việc học các đặc trưng ẩn của dữ liệu để tiến hành dự đoán gặp rất nhiều khó khăn; (3) - Một chỉ số tài chính (e.g. giá cổ phiếu của một công ty) không hoàn toàn phụ thuộc vào dữ liệu quá khứ mà còn phụ thuộc nhiều vào các tin tức, tình hình kinh tế, chính trị [15].

Đối với thách thức đầu tiên, các mô hình ensemble [cite something here] thường được sử dụng để hạn chế ảnh hưởng của sự biến đổi variance. Chúng tôi cũng tiếp cận bài toán theo hướng này nhưng ở mức cao hơn bằng cách sử dụng Meta-learning (ML) [8]. Phương pháp này tổng hợp hiệu quả tham số của các mô hình cục bộ, giúp giảm thiểu đáng kể variance loss.

Để khắc phục thách thức thứ hai, các nghiên cứu phần lớn sử dụng các đặc trưng rút trích được từ mô hình Long short-term memory neural network (LSTM) [11], Artificial neural network (ANN), và Convolution neural network (CNN) [14]. Cụ thể, trong năm 2022, 20% tổng số các bài

báo liên quan đến dự đoán chỉ số tài chính sử dụng LSTM, 20% sử dụng ANN và 6% sử dụng CNN [2]. Để tận dụng hết được các đặc trưng được rút trích bởi các mô hình nêu trên, chúng tôi đề xuất phương pháp kết hợp các đặc trưng này.

Đối với thách thức thứ ba, nghiên cứu [7] đưa ra giả thuyết rằng chỉ số tài chính của các công ty ngầm phản ánh tình hình kinh tế tài chính trên thị trường. Các nghiên cứu [1, 16] cũng chỉ ra sự phụ thuộc giữa chỉ số tài chính của một công ty nhất định và các chỉ số của các công ty khác. Điều này càng làm tăng tính đúng đắn của giả thuyết trong nghiên cứu [7]. Ngoài ra, chúng tôi cho rằng, các chỉ số tài chính còn có những phụ thuộc ngầm vào các thời điểm nhất định trong quá khứ (hidden-long-term dependency). Đối với cách tiếp cận truyền thống, người ta sử dụng một lượng dữ liệu quá khứ cố định (lookback window) để huấn luyện mô hình. Điều này gây một trở ngại lớn cho quá trình học vì các đặc trưng dài hạn theo thời gian sẽ bị quên. Mặt khác, ML chia nhỏ tập dữ liệu thành nhiều phần để học và tổng hợp hiệu quả các tham số học được nên có thể xử lý tốt thách thức này.

Cuối cùng, chúng tôi chứng minh tính ưu việt của thuật toán đề xuất bằng cách giải bài toán dự đoán xu hướng (tăng hoặc giảm) của tỉ giá ngoại hối và so sánh với mô hình SOTA hiện tại (NHITS [3]) trên hai loại dữ liệu: (1) - Dữ liệu tỉ giá của cặp tiền tệ USD/JPY; (2) - Dữ liệu tỷ giá của 60 cặp tiền tệ, cấu thành từ các nước Australia, Canada, Switzerland, Denmark, EU, United Kingdom, Hong Kong, Iceland, Japan, Norway, New Zealand, Singapore, Sweden, Turkey, United States, Mexico, China, South Africa. Các tập dữ liệu này được công khai trên Internet và có thể tải về dễ dàng. Bản cài đặt chính thức có thể xem tại [bỏ cái link github vào đây].

Tóm lại, đóng góp chính của chúng tôi như sau:

- Tổng hợp hiệu quả tham số mô hình: Sử dụng ML thay cho các mô hình ensemble truyền thống trong việc tổng hợp kết quả từ các mô hình học máy.
- Kết hợp đặc trưng: Rút trích đặc trưng bằng cách kết hợp các đặc trưng của LSTM và CNN.
- Hidden-long-term dependency: Chứng minh thực

* mwklng2309@icloud.com

† ichibha@icloud.com

‡ rmaezono@mac.com

nghiệm rằng một chỉ số tài chính không chỉ phụ thuộc vào các chỉ số tài chính khác mà còn có các phụ thuộc ẩn với chính nó.

- Experiment: Thực nghiệm trên các bộ dữ liệu về tỉ giá hối đoái và so sánh với mô hình SOTA NHITS để chứng minh tính hiệu quả của phương pháp đề xuất.

II. RELATED WORK

A. LSTM & CNN in feature extraction

Như đã đề cập, LSTM là mạng neural rất phổ biến trong việc handle các bài toán liên quan đến dự đoán chỉ số tài chính nói riêng cũng như các bài toán trên dữ liệu time-series nói chung. LSTM được sử dụng phổ biến như vậy bởi vì nó xử lý tốt vấn đề vanishing gradient và có thể khai thác các mối quan hệ phi tuyến trong dữ liệu. Thật vậy, bằng cách duy trì cell-state trong mỗi iteration, LSTM có thể chống lại một cách hiệu quả vấn đề vanishing gradient, từ đó bảo toàn khả năng capture các phụ thuộc dài hạn [4]. LSTM cũng thực hiện rút trích đặc trưng với các hàm kích hoạt phi tuyến, giúp học được các tham số mô hình có thể capture tính phi tuyến của dữ liệu [9]. Hai yếu tố nêu trên khiến cho LSTM trở thành lựa chọn đầu tiên được nghĩ đến khi giải quyết các bài toán trên dữ liệu time-series.

CNN được sử dụng rất nhiều trong các tác vụ xử lý hình ảnh [17, 18] bởi khả năng tổng hợp các quan hệ cục bộ. Không chỉ vậy, CNN còn được dùng rất nhiều trong các tác vụ xử lý dữ liệu time-series như speech recognition [5], natural language processing [19]. Điều đó chứng tỏ được khả năng của CNN trong việc khám phá mối quan hệ thời gian giữa các mẫu dữ liệu. Mặc dù vậy, CNN lại rất ít được dùng trong các tác vụ dự đoán chỉ số tài chính. Trong nghiên cứu này, chúng tôi tận dụng khả năng rút trích đặc trưng cục bộ tuyệt vời của CNN để tích hợp thêm thông tin ẩn vào quá trình huấn luyện của mô hình.

B. Meta-learning

Các thuật toán Meta-learning (ML), điển hình là Model-agnostic meta-learning (MAML) [8] được biết đến với khả năng huấn luyện một mô hình có tính tổng quát cao, thích ứng nhanh trên tập dữ liệu mới thông qua một lượng nhỏ dữ liệu và số bước huấn luyện [12, 20]. Với khả năng này, ML được sử dụng rất nhiều trong việc cải thiện tốc độ đáp ứng của mô hình trên dữ liệu như cá nhân hóa mô hình học [6], giải quyết vấn đề domain shift [insert citation].

Một thuật toán ML cơ bản sẽ được học trên nhiều tác vụ t rút ra từ cùng một phân phối tác vụ \mathcal{T} [12]. Dữ liệu của mỗi tác vụ được chia thành tập support $\mathcal{D}_t^{support}$ (thường có kích thước nhỏ, khoảng 20%) và tập query

\mathcal{D}_t^{query} . Trong quá trình học, hai bước tối ưu inner và outer optimization được perform đan xen. Inner optimization cố gắng tìm ra một bộ tham số tối ưu θ_t^* cho từng mô hình học máy trên tập support của mỗi tác vụ bằng phương trình 1.

$$\theta_t^* = \theta_t(\phi) = \arg \min_{\theta} \mathcal{L}_t(\phi, \mathcal{D}_t^{support}) \quad (1)$$

Trong đó, ϕ là kết quả của quá trình outer optimization, đóng vai trò là giá trị khởi tạo của θ_t . \mathcal{L}_t là hàm loss của mô hình trên tác vụ t .

Sau đó, thuật toán sử dụng các bộ tham số tối ưu θ_t^* để perform trên tập query tương ứng. Loss của toàn bộ mô hình sau đó được tổng hợp để thực hiện quá trình outer optimization như phương trình 2.

$$\begin{aligned} \phi^* &= \arg \min_{\phi} \sum_t \mathcal{L}_t[\theta_t^*, \mathcal{D}_t^{query}] \\ &= \arg \min_{\phi} \sum_t \mathcal{L}_t[\theta_t(\phi), \mathcal{D}_t^{query}] \end{aligned} \quad (2)$$

Trong inference phase, giá trị khởi tạo cho tham số của mô hình được gán bằng ϕ^* . Mô hình sau đó được huấn luyện nhanh trên tập support sau đó perform trên tập query. Kết quả trên tập query chính là kết quả của mô hình.

Bằng hình thức huấn luyện trên, mô hình ϕ^* sẽ có mức tổng quát hóa cao trên các tác vụ khác nhau, có thể nhanh chóng đáp ứng một tác vụ mới chỉ sau một vài bước huấn luyện. Tận dụng ưu điểm này, chúng tôi

Các mô hình hybrid ensemble vốn được sử dụng rất nhiều trong các bài toán xử lý time-series và được chứng minh thực nghiệm là có độ chính xác cao hơn so với các mô hình handle time-series data tiêu chuẩn vì có thể tổng hợp được sức mạnh của nhiều mô hình [2]. Tuy vậy, các hình thức tổng hợp của ensemble model hiện nay vẫn còn rất cứng nhắc vì chỉ có thể tổng hợp dựa trên kết quả cuối (đối với bagging models) và kết quả gần cuối (đối với stacking models). Dưới góc nhìn của ensemble model, có thể coi phương trình 2 là một phương pháp tổng hợp hiệu quả các sub-model, giúp tận dụng khả năng rút trích đặc trưng của từng mô hình. Nói cách khác, mô hình sau khi tổng hợp có thể rút trích đặc trưng ở mức sâu hơn, cải thiện đáng kể khả năng dự đoán so với các mô hình ensemble truyền thống.

C. Neural Hierarchical Interpolation for Time Series (NHITS)

NHITS được thiết kế để hướng đến việc dự đoán các long-horizon time-series data. Theo nghiên cứu [3], cấu trúc của NHITS bao gồm nhiều stack liên tiếp nhau. Mỗi stack bao gồm nhiều block nối tiếp nhau. Tại mỗi block, dữ liệu lịch sử được sử dụng để dự đoán dữ liệu tương

lai và dữ liệu quá khứ. Cụ thể, tại block l , với L mẫu dữ liệu quá khứ ($\mathbf{y}_{t-L:t,l-1}$), các đặc trưng sẽ được rút trích như sau (theo [3]):

$$\mathbf{y}_{t-L:t,l}^{(p)} = \text{Pooling}(\mathbf{y}_{t-L:t,l-1}) \quad (3)$$

$$\theta_l^b = \text{FullyConnected}^b(\mathbf{y}_{t-L:t,l}^{(p)}) \quad (4)$$

$$\theta_l^f = \text{FullyConnected}^f(\mathbf{y}_{t-L:t,l}^{(p)}) \quad (5)$$

$$\hat{\mathbf{y}}_{t-L:t,l} = g(\theta_l^b) \quad (6)$$

$$\hat{\mathbf{y}}_{t+1:t+H,l} = g(\theta_l^f) \quad (7)$$

Trong đó, **FullyConnected** là các lớp multi-layer perception (MLP) xếp chồng với hàm kích hoạt phi tuyến. θ_l^f, θ_l^b là các hệ số nội suy forecast và backcast, được dùng để tổng hợp các giá trị đầu ra của block l bằng hàm nội suy $g(\cdot)$. Đầu ra của block l là giá trị forecast $\hat{\mathbf{y}}_{t+1:t+H,l}$ và giá trị backcast $\hat{\mathbf{y}}_{t-L:t,l}$. Input của block $l+1$ được tính theo phương trình 8.

$$\mathbf{y}_{t-L:t,l+1} = \mathbf{y}_{t-L:t,l} - \hat{\mathbf{y}}_{t+1:t+H,l} \quad (8)$$

Tổng hợp các giá trị forecast của các block, ta được giá trị forecast của một stack. Backcast của block cuối cùng của một stack chính là đầu vào cho stack tiếp theo. Cuối cùng, tổng hợp giá trị forecast của các stack, ta được giá trị forecast dự đoán của toàn mạng.

Bằng cách xếp chồng các stack, stack sau nhận vào phần dư của stack trước, kiến trúc trên được kỳ vọng là sẽ phân rã dữ liệu thành các frequency bands khác nhau. Trong thực tế, NHITS perform rất tốt đối với các bộ dữ liệu có tính chu kỳ cao. Tuy nhiên, tính chu kỳ của dữ liệu foreign exchange là rất thấp, thậm chí không có, gây ra khó khăn rất lớn cho NHITS. Bằng cách xếp chồng các stack, stack sau nhận vào phần dư của stack trước, kiến trúc trên được kỳ vọng là sẽ phân rã dữ liệu thành các frequency bands khác nhau. Trong thực tế, NHITS perform rất tốt đối với các bộ dữ liệu có tính chu kỳ cao. Tuy nhiên, tính chu kỳ của dữ liệu foreign exchange là rất thấp, thậm chí không có, gây ra khó khăn rất lớn cho NHITS.

III. METHODOLOGY

Fixed-node DMC

Fixed-node DMC

$$\left[\sum_j \frac{1}{2} \nabla_j^2 + V_{\text{MB}}(\vec{R}) \right] \Psi(\vec{R}) = E \Psi(\vec{R}) . \quad (9)$$

Fixed-node DMC

$$\left[\frac{1}{2} \nabla_j^2 + v_{\text{ext}}(\vec{r}_j) + v_{\text{XC}}(\vec{r}_j) \right] \psi(\vec{r}_j) = \varepsilon \psi(\vec{r}_j) . \quad (10)$$

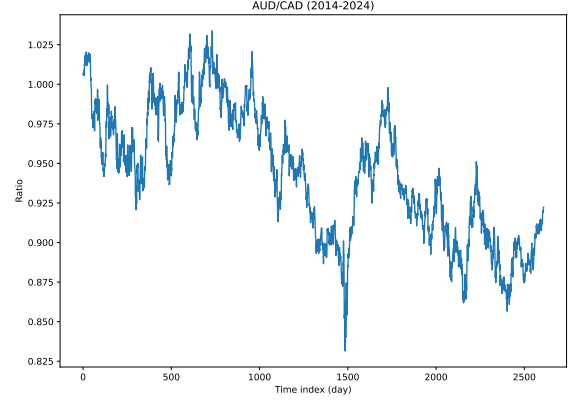


FIG. 1. Your figure caption here.

Fixed-node DMC

Fixed-node DMC

$$E[\Psi_T] = \frac{\int d\vec{R} \cdot \Psi_T^*(\vec{R}) \hat{H} \Psi_T(\vec{R})}{\int d\vec{R} \cdot \Psi_T^*(\vec{R}) \Psi_T(\vec{R})}, \quad (11)$$

Fixed-node DMC

Fixed-node DMC

Fixed-node DMC

IV. EXPERIMENT

A. Density functional theory

Density functional theory

B. Fixed-node DMCTEXT

Fixed-node DMCTEXT

Fixed-node DMCTEXT

Fixed-node DMCTEXT

V. RESULTS AND DISCUSSION

Results and discussion

Results and discussion

Results and discussion

VI. CONCLUSION

Conclusion

VII. ACKNOWLEDGEMENTS

The work was carried out within the state assignment of Ministry of Science and Higher Education of the Russian Federation (No. AAAA-A18-118020190095-4, topic “Quantum”). G.I.P is grateful for financial supports from JST SPRING, Grant Number JPMJSP2102. K. H. is grateful for financial support from MEXT-KAKENHI (JP19K05029, JP21K03400, JP21H01998, and JP22H02170). R.M. is grateful for financial supports from MEXT-KAKENHI (22H05146, 21K03400 and 19H04692).

VIII. AUTHOR CONTRIBUTIONS

All authors contributed to conceiving the idea. T.I., A.V.U, and G.I.P performed the calculations. T.I. made the all figures. R.M and S.V.S supervised the work. All authors contributed to the discussion and writing of the paper.

IX. DATA AVAILABILITY STATEMENT

The datasets used and/or analyzed during the current study available from the corresponding authors on reasonable request.

-
- [1] Andraw W. Lo, A.C.M., 1990. When are contrarian profits due to stock market overreaction? The review of financial studies 3, 175–205.
 - [2] Ayitey Junior, M., Appiahene, P., Appiah, O., Bombie, C.N., 2023. Forex market forecasting using machine learning: Systematic literature review and meta-analysis. *Journal of Big Data* 10, 9.
 - [3] Challu, C., Olivares, K.G., Oreshkin, B.N., Ramirez, F.G., Canseco, M.M., Dubrawski, A., 2023. Nhits: Neural hierarchical interpolation for time series forecasting, in: *Proceedings of the AAAI conference on artificial intelligence*, pp. 6989–6997.
 - [4] Cheng, L.C., Huang, Y.H., Wu, M.E., 2018. Applied attention-based lstm neural networks in stock prediction, in: *2018 IEEE International Conference on Big Data (Big Data)*, IEEE. pp. 4716–4718.
 - [5] Dua, S., Kumar, S.S., Albagory, Y., Ramalingam, R., Dumka, A., Singh, R., Rashid, M., Gehlot, A., Alshamrani, S.S., AlGhamdi, A.S., 2022. Developing a speech recognition system for recognizing tonal speech signals using a convolutional neural network. *Applied Sciences* 12, 6223.
 - [6] Fallah, A., Mokhtari, A., Ozdaglar, A., 2020. Personalized federated learning: A meta-learning approach. *arXiv preprint arXiv:2002.07948*.
 - [7] Fama, E.F., 1970. Efficient capital markets: A review of theory and empirical work. *Journal of finance* 25, 383–417.
 - [8] Finn, C., Abbeel, P., Levine, S., 2017. Model-agnostic meta-learning for fast adaptation of deep networks, in: *International conference on machine learning*, PMLR. pp. 1126–1135.
 - [9] He, T., Droppo, J., 2016. Exploiting lstm structure in deep neural networks for speech recognition, in: *2016 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, IEEE. pp. 5445–5449.
 - [10] Heryadi, Y., Wibowo, A., et al., 2021. Foreign exchange prediction using machine learning approach: A pilot study, in: *2021 4th International Conference on Information and Communications Technology (ICOIACT)*, IEEE. pp. 239–242.
 - [11] Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. *Neural computation* 9, 1735–1780.
 - [12] Hospedales, T., Antoniou, A., Micaelli, P., Storkey, A., 2021. Meta-learning in neural networks: A survey. *IEEE transactions on pattern analysis and machine intelligence* 44, 5149–5169.
 - [13] Islam, M.S., Hossain, E., 2021. Foreign exchange currency rate prediction using a gru-lstm hybrid network. *Soft Computing Letters* 3, 100009.
 - [14] LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., Jackel, L., 1989. Handwritten digit recognition with a back-propagation network. *Advances in neural information processing systems* 2.
 - [15] Li, C., Song, D., Tao, D., 2019. Multi-task recurrent neural networks and higher-order markov random fields for stock price movement prediction: Multi-task rnn and higher-order mrfs for stock price classification, in: *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 1141–1151.
 - [16] Mech, T.S., 1993. Portfolio return autocorrelation. *Journal of Financial Economics* 34, 307–344.
 - [17] Naranjo-Torres, J., Mora, M., Hernández-García, R., Barrientos, R.J., Fredes, C., Valenzuela, A., 2020. A review of convolutional neural network applied to fruit image processing. *Applied Sciences* 10, 3443.
 - [18] Sharma, N., Jain, V., Mishra, A., 2018. An analysis of convolutional neural networks for image classification. *Procedia computer science* 132, 377–384.
 - [19] Varshitha, K.S., Kumari, C.G., Hasvitha, M., Fiza, S., Amarendra, K., Rachapudi, V., 2023. Natural language processing using convolutional neural network, in: *2023 7th International Conference on Computing Methodologies and Communication (ICCMC)*, IEEE. pp. 362–367.
 - [20] Vettoruzzo, A., Bouguelia, M.R., Vanschoren, J., Rognvaldsson, T., Santosh, K., 2024. Advances and challenges in meta-learning: A technical review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.