

ĐẠI HỌC QUỐC GIA TP.HCM
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN



MÔN: KHAI THÁC DỮ LIỆU ĐỒ THỊ
BÁO CÁO CUỐI KÌ

BIASED TEMPORAL CONVOLUTION GRAPH NETWORK FOR
TIME SERIES FORECASTING WITH MISSING VALUES

(MẠNG TÍCH CHẬP ĐỒ THỊ THỜI GIAN CÓ HƯỚNG LỆCH DÀNH
CHO VIỆC DỰ ĐOÁN CHUỖI THỜI GIAN CÓ DỮ LIỆU BỊ THIẾU)

Họ và tên: Nguyễn Đức Tuấn Đạt - 21127590
Giáo viên phụ trách: Lê Ngọc Thành - Lê Nhật Nam

TP. Hồ Chí Minh, năm 2024

Mục lục

1	Giới thiệu	2
1.1	Nêu vấn đề	2
1.2	Ý tưởng giải quyết	2
1.3	Phương pháp giải quyết	2
1.4	Ý nghĩa thực tiễn	3
2	Mục đích nghiên cứu	4
2.1	Tại sao cần thực hiện đề tài này?	4
2.2	Đề tài mang lại được điều gì?	5
2.3	Ảnh hưởng và ý nghĩa có thể có của kết quả đối với vấn đề đã được đặt ra nói riêng và toàn bộ hướng nghiên cứu nói chung?	5
3	Mục tiêu nghiên cứu chính của đề tài	7
3.1	Mục tiêu nghiên cứu chính	7
3.2	Nội dung nghiên cứu chính	7
4	Tổng quan phương pháp được đề xuất	9
4.1	Các khái niệm cơ bản	9
4.1.1	Hiểu về dữ liệu đầu vào	9
4.1.2	Hiểu về dự đoán chuỗi thời gian đa biến với dữ liệu bị thiếu	11
4.2	Chi tiết phương pháp chính	11
4.2.1	Luồng làm việc của BiTGraph - Framework giải quyết bài toán	11
4.2.2	Multi-Scale Instance PartialTCN Module	12
4.2.3	Biased GCN Module	15
4.2.4	Kiến trúc phân cấp - Hierarchical Architecture	18
4.3	Nghiên cứu cắt bỏ	18
5	Kết quả cuối cùng	20

5.1	Cơ sở so sánh	20
5.2	Thiết lập đầu vào	21
5.3	Kết quả tổng quan	22
5.4	Nhận xét	22
6	Kết luận	26
7	Thông tin thêm	27
7.1	Độ nhạy cảm của HYPERPARAMETER	27
7.2	Vai trò của β	28
7.3	Trực quan kết quả giữa BiTGraph với MTGNN	28
7.4	Độ phức tạp của mô hình	29
8	Nguồn tham khảo	30

Giới thiệu

1.1. Nêu vấn đề

- Dự báo chuỗi thời gian đa biến chiếm 1 vị trí rất quan trọng trong rất nhiều lĩnh vực trải dài từ khí tượng học, quản lý giao thông đến kế hoạch hóa kinh tế.
- Trải qua nhiều thập kỉ, rất nhiều sự nỗ lực đã được bỏ ra để hướng tới độ chính xác và đáng tin cậy trong việc phát triển những phương pháp dự đoán dưới giả định rằng dữ liệu đầu vào ổn định và không bị lỗi.
- Tuy nhiên, dữ liệu về chuỗi thời gian đến từ thực tế lại thường bị thiếu do vấn đề trực tiếp kĩ thuật hoặc chi phí thu thập dữ liệu quá cao, dẫn đến sự ảnh hưởng nghiêm trọng trong hiệu suất của các phương pháp hiện có.

1.2. Ý tưởng giải quyết

- Nhóm tác giả đã đưa ra một nghiên cứu “Mạng lưới đồ thị tích chập thời gian có thiên vị” có khả năng đồng thời nắm bắt những phụ thuộc theo thời gian và cấu trúc không gian.
- Nói đúng hơn là họ bổ sung những thiên vị vào 2 khối được thiết kế cẩn thận - the Multi-Scale Instance PartialTCN and Biased GCN - để tính toán những mẫu dữ liệu bị thiếu.
- Multi-Scale Instance PartialTCN để mô hình hiệu quả "phụ thuộc thời gian" bị gián đoạn do thiếu dữ liệu và đưa ra Biased GCN để lan truyền thông tin giữa các phần tử bằng việc xây dựng 1 đồ thị có trọng số lệch có tính đến các mẫu thiếu hụt.

1.3. Phương pháp giải quyết

- Multi-Scale Instance PartialTCN thực hiện tích chập thời gian cục bộ không phụ thuộc vào phần tử để nắm bắt các phụ thuộc theo thời gian bên trong từng phần tử, ngay cả khi dữ liệu bị ảnh hưởng bởi dữ liệu thiếu.

- Thêm vào đó, Biased GCN còn khám phá được cấu trúc không gian dữ liệu bằng cách xây dựng một đồ thị có trọng số lệch để tính toán các mẫu thiếu hụt.
- Bên cạnh đó, họ sẽ tích hợp 2 mô-đun dựa theo 1 kiến trúc phân cấp(hierarchical architecture), nơi mà các mẫu bị thiếu dữ liệu sẽ được cập nhật dần dần(progressively updated) theo cả chiều thời gian và không gian để tối đa hóa việc lan truyền thông tin (maximize information propagation) và giảm thiểu tác động của những dữ liệu bị thiếu(minimizes the impact of missing values).

1.4. Ý nghĩa thực tiễn

- BiTGraph đạt đến 9.93% về mặt hiệu quả so với hơn những phương pháp dự báo trước đó trong ngữ cảnh nhiều dữ liệu bị thiếu, được xác nhận trong 5 mẫu dữ liệu chuẩn thực tế.

Mục đích nghiên cứu

2.1. Tại sao cần thực hiện đề tài này?

- Như trước đó chúng ta đã đề cập đến tầm quan trọng của dự báo chuỗi thời gian đa biến, nhưng chủ yếu thì các phương pháp trước đó chỉ hoạt động tốt chủ yếu trên những bộ dữ liệu đầy đủ và không bị thiếu hay sai sót gì.
- Tuy nhiên, Dữ liệu về chuỗi thời gian đến từ thực tế thường bị thiếu do vấn đề trục trặc kĩ thuật hoặc chi phí thu thập dữ liệu quá cao, dẫn đến sự ảnh hưởng nghiêm trọng trong hiệu suất của các phương pháp hiện có.
- Việc sử dụng những phương pháp đơn giản để xử lí dữ liệu thiếu thì không thể tránh khỏi những tích lũy về lỗi và dẫn đến những giải pháp không tối ưu, cụ thể chi tiết như sau:
 - Các phương pháp thống kê (statistical method) bao gồm ARIMA(Autoregressive Integrated Moving Average) và VAR(Vector Autoregression). Đây là những phương pháp thuở ban đầu trong việc xây dựng mô hình tự hồi quy để bắt được những phụ thuộc về thời gian (temporal dependencies). Tuy nhiên, giả định phụ thuộc tuyến tính của chúng lại thường dẫn đến sự nghèo nàn về mặt hiệu quả trong thực tế.
 - Những phương pháp dựa vào RNNs tận dụng sức mạnh biểu đạt (expressive power) của mạng nơ ron vào thời điểm hiện tại đã nói lỏng đi được những giả định phụ thuộc tuyến tính. Gần đây nhất, nhiều phương pháp đa dạng dựa trên Transformer đã được đề xuất để tận dụng những vùng tiếp nhận rộng (wide-range receptive fields) của cơ chế chú ý cho dự đoán dài hạn.
 - BRITS đề xuất sử dụng một lớp kết nối dày (dense connection layer) để học về sự tương quan giữa các phần tử(instance pair), nhưng kết quả chỉ ra rằng độ phức tạp của mô hình(model complexity) rất cao.
 - Để giảm thiểu độ phức tạp bậc hai của cơ chế chú ý truyền thống, Informer, Pyraformer, Autoformer và FEDformer đã được đề xuất lần lượt.

- Non-stationary Transformer nhằm để việc tập trung cải thiện cơ chế tập trung để thích ứng với đặc tính không tĩnh của chuỗi dữ liệu thời gian.
- PatchTST khám phá ra thiết kế theo miếng vá (patch) và kênh độc lập (channel-independence).
- DCRNN, AGCRN, MTGNN, GTS, SAGDFN là những mô hình cấu trúc không gian với mạng đồ thị nơ-ron. Thêm vào đó, CoST và TS2Vec hướng tới dự đoán chuỗi thời gian từ khía cạnh học tự quan sát (self-supervised learning)
- Mô hình chuỗi thời gian với dữ liệu thiếu do lỗi về thiết bị, lỗi về giao tiếp hoặc khó khăn trong thu thập dữ liệu, dữ liệu chuỗi thời gian thu thập theo thế giới thực thường không hoàn chỉnh và chỉ được quan sát 1 phần. Để làm đầy được những lỗ hổng này, rất nhiều phương pháp imputation thời gian thực - BRITS, GRIN, CSDI, SPIN, GRIN, TIDER - đã được trình bày trong cộng đồng học máy. Để giải quyết được vấn đề chuỗi thời gian bị quan sát 1 phần, 1 cách có khả năng để xây dựng được những mô hình dự đoán với những kết quả được suy ra từ phương pháp xử lý dữ liệu khuyết (imputation methods). Tuy nhiên, phương pháp imputation này lại tách biệt hoàn toàn so với việc dự đoán trong 2 bước quá trình, và điều này khiến cho lỗi tích lũy ngày càng ảnh hưởng nghiêm trọng đến hiệu quả dự đoán.

2.2. Đề tài mang lại được điều gì?

- Đề tài này đã cung cấp ra hướng tiếp cận khác trong việc Dự báo chuỗi thời gian đa biến, đặc biệt là trong trường hợp bộ dữ liệu đầu vào không hoàn chỉnh và bị thiếu do những tác nhân khác nhau đến từ thực tế.
- Bài báo giới thiệu Multi-Scale Instance PartialTCN để mô hình hóa các phụ thuộc thời gian và Biased GCN để nắm bắt cấu trúc không gian, cả hai đều được thiết kế để xử lý hiệu quả các mẫu dữ liệu bị thiếu. Để từ đó, BiTGraph cho thấy hiệu suất vượt trội so với các phương pháp hiện có trên các tập dữ liệu thực tế, đặc biệt là khi tỷ lệ dữ liệu thiếu cao.

2.3. Ảnh hưởng và ý nghĩa có thể có của kết quả đối với vấn đề đã được đặt ra nói riêng và toàn bộ hướng nghiên cứu nói chung?

- Đối với vấn đề được đặt ra:

- Kết quả thực nghiệm cho thấy BiTGraph có khả năng cải thiện đáng kể độ chính xác của các mô hình dự báo chuỗi thời gian đa biến trong trường hợp dữ liệu bị thiếu.
 - Mô hình này giải quyết trực tiếp một vấn đề thường gặp trong thực tế, nơi dữ liệu chuỗi thời gian thường không đầy đủ.
- Đối với hướng nghiên cứu:
 - Bài báo đề xuất một cách tiếp cận mới để xử lý dữ liệu bị thiếu trong dự báo chuỗi thời gian, thay vì chỉ dựa vào các phương pháp nội suy truyền thống.
 - Mô hình và các mô-đun được đề xuất có thể được áp dụng và mở rộng cho nhiều ứng dụng khác nhau liên quan đến dữ liệu chuỗi thời gian bị thiếu ngoài những lĩnh vực được đề cập trước đó như là: chăm sóc sức khỏe, tài chính, nông nghiệp...

Mục tiêu nghiên cứu chính của đề tài

3.1. Mục tiêu nghiên cứu chính

Phát triển một mô hình dự báo chuỗi thời gian đa biến có khả năng xử lý hiệu quả các giá trị bị thiếu, thay vì chỉ dựa vào các phương pháp nội suy trước đó. Mô hình này cần phải nắm bắt được cả hai yếu tố quan trọng trong chuỗi thời gian đa biến: sự phụ thuộc theo thời gian (trong cùng một chuỗi) và cấu trúc không gian (giữa các chuỗi khác nhau).

3.2. Nội dung nghiên cứu chính

- Đối tượng nghiên cứu: Dữ liệu chuỗi thời gian đa biến, đặc biệt là các chuỗi thời gian có chứa giá trị bị thiếu.
- Thực thể liên quan:
 - Mô hình BiTGraph (Biased Temporal Convolution Graph Network): Mô hình chính được đề xuất trong bài báo, kết hợp các mô-đun Multi-Scale Instance PartialTCN và Biased GCN.
 - Mô-đun Multi-Scale Instance PartialTCN: Mô-đun này tập trung vào việc nắm bắt các phụ thuộc thời gian trong từng chuỗi con, có khả năng xử lý các giá trị bị thiếu.
 - Mô-đun Biased GCN: Mô-đun này khai thác cấu trúc không gian giữa các chuỗi con bằng cách xây dựng một đồ thị có trọng số (biased graph) để phản ánh các mẫu dữ liệu bị thiếu.
- Tập dữ liệu: Nghiên cứu có sử dụng 5 tập dữ liệu thực tế từ các lĩnh vực khác nhau (Metr-LA, Electricity, PEMS, ETTh1, và BeijingAir) để đánh giá hiệu suất của mô hình.
- Giới hạn hoặc ràng buộc:
 - Mô hình giả định rằng các giá trị bị thiếu là ngẫu nhiên và không có mẫu cụ thể.

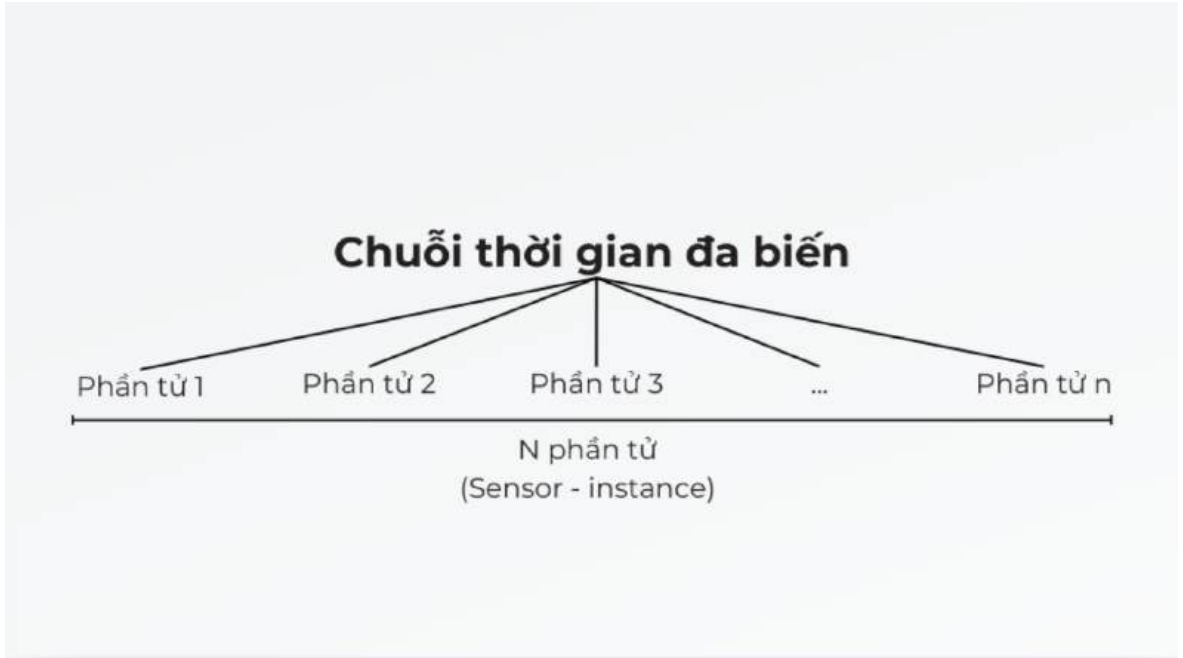
- Hiệu suất của mô hình có thể bị ảnh hưởng bởi các yếu tố như tỷ lệ giá trị bị thiếu, độ phức tạp của chuỗi thời gian, và sự lựa chọn các siêu tham số.

Tổng quan phương pháp được đề xuất

4.1. Các khái niệm cơ bản

4.1.1. Hiểu về dữ liệu đầu vào

- Dữ liệu chuỗi thời gian đa biến được tạo ra bởi một tập hợp N phần tử (N instances - sensors) trong một khoảng thời gian nhất định.

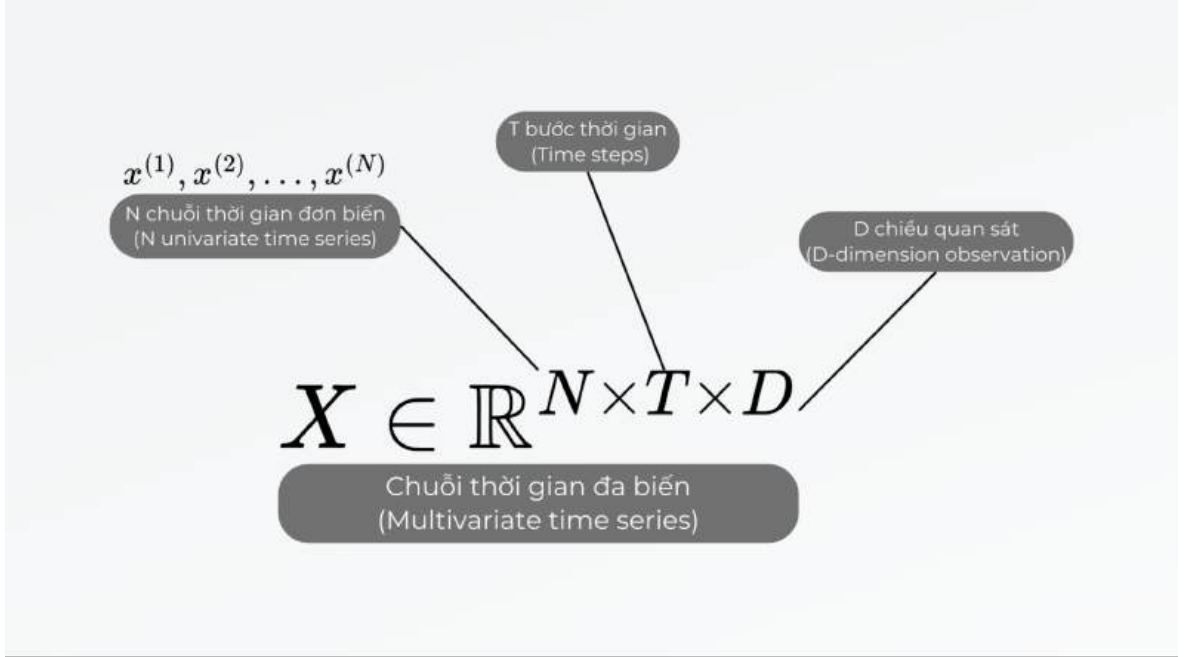


Hình 4.1: Minh họa cho các phần tử của dữ liệu chuỗi thời gian đa biến.

- Trong bài báo này, chúng ta sẽ làm việc với chuỗi thời gian đa biến (multivariate time series) $X \in \mathbb{R}^{N \times T \times D}$ bao gồm N chuỗi thời gian đơn biến (univariate time series) $x^{(1)}, x^{(2)}, \dots, x^{(N)}$ được thu thập thông qua T bước thời gian (T time steps) với D chiều quan sát (D -dimension observation).
- Trong thực tế, lỗi khi thu thập dữ liệu luôn có khả năng xảy ra nên X có thể có tồn tại những giá trị bị lỗi (hay còn gọi là bị thiếu), và chúng ta sẽ sử dụng 1 ma trận mặt nạ (mask matrix) $M \in \mathbb{R}^{N \times T}$ để đại diện cho những mẫu bị lỗi này, có thể được biểu diễn

với công thức dưới đây:

$$M_{nt} = \begin{cases} 1, & \text{nếu } X_{nt} \text{ được quan sát,} \\ 0, & \text{ngược lại,} \end{cases} \quad (4.1)$$



Hình 4.2: Ý nghĩa của các chiều trong dữ chuỗi thời gian đa biến được sử dụng trong bài báo.

- Ngoài ra, X_{nt} sẽ thể hiện giá trị của phần tử (hoặc kênh) thứ n tại bước t , chúng ta cũng có thể sử dụng $x_t^{(n)}$, để biểu hiện là cùng 1 đầu vào.
- Tương tự như vậy, ta sẽ dùng $m^{(n)} \in \mathbb{R}^T$ để biểu diễn hàng thứ n của ma trận mặt nạ M , hoặc có thể trình bày bằng $M_{nt} - m_t^{(n)}$ với ý nghĩa là đại diện cho phần tử tại hàng thứ n , cột thứ t của M .
- Thêm vào đó, kí hiệu cắt lát $x_{t-H:t} \in \mathbb{R}^{H \times D}$ hay $X_{t-H:t} \in \mathbb{R}^{N \times H \times D}$ sẽ biểu diễn những giá trị tại ô của sổ thời gian có kích thước H tính từ bước $t - h$ đến $t - 1$, tức là khoảng thời gian $[t - H, t)$.
- Nhóm nghiên cứu đã chọn năm bộ dữ liệu dự báo chuỗi thời gian được sử dụng phổ biến nhất: Metr-LA, Electricity, PEMS, ETTh1 và BeijingAir, các thống kê của chúng được tóm tắt trong Bảng 4.1. Năm bộ dữ liệu được thu thập từ các miền khác nhau và bao

	Metr-LA	Electricity	PEMS	ETTh1	BeijingAir
#Samples (T)	34272	26304	52116	17420	8759
#Instances (N)	207	321	325	7	36
Frequency	5 min	1 h	5 min	1 h	1 h
Mean	53.72	2538.79	62.62	4.58	72.01
Variance	410.53	2.26×10^8	92.05	42.68	79.07

Bảng 4.1: Mô tả dữ liệu

gồm các phạm vi cường độ, tần số lấy mẫu và thống kê đa dạng. Họ loại bỏ ngẫu nhiên dữ liệu theo tỷ lệ thiếu r dao động từ 0,1 đến 0,8, bao gồm 0,1, 0,2, 0,4, 0,6 và 0,8.

4.1.2. Hiểu về dự đoán chuỗi thời gian đa biến với dữ liệu bị thiếu

- Với chuỗi thời gian đa biến được quan sát 1 phần (tức là giá trị trong chuỗi thời gian bị lỗi) X và ma trận mặt nạ M tương ứng, để dự đoán chuỗi thời gian đa biến khi tồn tại những giá trị bị thiếu thì ta chúng ta phải nhắm đến việc xây dựng 1 mô hình dự đoán ϕ để dự đoán giá trị bước F tiếp theo $Y = X_{t:t+F}$ bằng cách lấy quan sát trong quá khứ $X_{t-H:t}$ như những đầu vào và mặt nạ của nó $M_{t-H:t}$ có thể viết là $\hat{Y} = \phi(X_{t-H:t}, M_{t-H:t})$.
- Trong giai đoạn huấn luyện, chúng ta chỉ sử dụng những giá trị quan sát được để chứng cấp những tín hiệu học (learning signals). Nói theo 1 cách chính thức hơn đó là hàm mất mát \mathcal{L} của mô hình có thể được mô tả như sau:

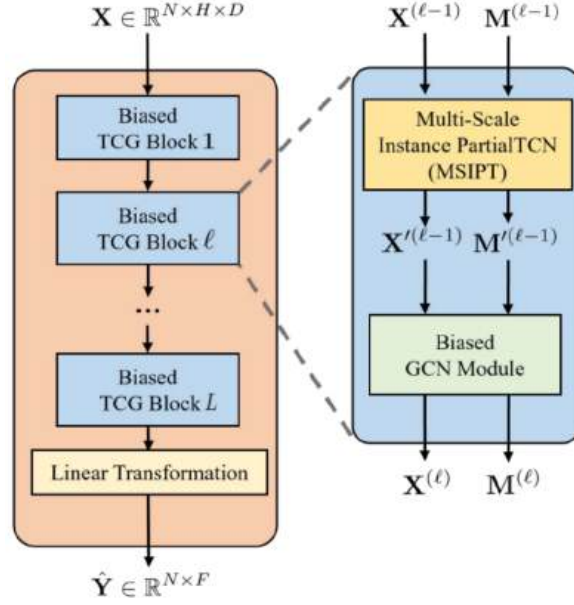
$$\mathcal{L}(Y, \hat{Y}, M_{t:t+F}) = \frac{\sum_{n=1}^N \sum_{\mathcal{T}=1}^{t+F-1} m_{\mathcal{T}}^{(n)} |\hat{y}_{\mathcal{T}}^{(n)} - y_{\mathcal{T}}^{(n)}|}{\sum_{n=1}^N \sum_{\mathcal{T}=1}^{t+F-1} m_{\mathcal{T}}^{(n)}} \quad (4.2)$$

Hàm trên để đo lường sai số tuyệt đối trung bình giữa những giá trị dự đoán và giá trị thực.

4.2. Chi tiết phương pháp chính

4.2.1. Luồng làm việc của BiTGraph - Framework giải quyết bài toán

- Khung vận hành của BiTGraph của chúng ta được trình bày ở Hình 4.3. Framework này bao gồm L khối giống hệt nhau, được gọi là Biased TCGBlock (Khối tích chập thời gian đồ thị có hướng lệch), đây là khối xây dựng cơ bản của phương pháp của chúng ta.
- Biased TCGBlock bao gồm 2 mô-đun chính: Multi-scale Instance PartialTCN và Biased GCN. 2 mô-đun này chịu trách nhiệm cho hợp nhất thông tin dựa trên chiều thời gian



Hình 4.3: Khung vận hành của BiTGraph

và chiều không gian.

- Ngược lại với những phương pháp dự đoán chuỗi thời gian đã tồn tại, nhóm tác giả đã cân nhắc và sử dụng bias để tính toán những mẫu thiếu trong thiết kế mô hình và cập nhật dần các mẫu thiếu như theo quá trình khuếch tán thông tin.
- Như Hình 4.3 có biểu diễn, khối thứ ℓ lấy những đầu vào $X^{(\ell-1)} \in \mathbb{R}^{N \times H \times D^{(\ell-1)}}$ và những mẫu bị thiếu $M^{\ell-1} \in \mathbb{R}^{N \times H}$, và nó đưa ra 2 tensors được biến đổi $X^{(\ell)} \in \mathbb{N} \times \mathbb{H} \times \mathbb{D}^{(\ell)}$ và $M^{(\ell)} \in \mathbb{R}^{N \times H}$, trong đó $D^{(\ell)}$ là chiều kích thước đặc trưng của khối thứ ℓ .

4.2.2. Multi-Scale Instance PartialTCN Module

- Nhóm tác giả đã chọn chọn Mạng tích chập thời gian (TCN) làm thành phần cốt lõi để nắm bắt sự biến đổi theo thời gian với 2 lí do chính:

- Nó đã chứng minh khả năng mô hình hóa chuỗi tốt hơn khi mà so sánh với RNNs trong nhiều tác vụ khác nhau.

Để có thể lý giải cho điều này, chúng ta sẽ tìm hiểu kĩ thông qua bài báo "An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling" [2] và đây là những gì mà em đúc rút được: Các tác giả đã thực hiện một đánh giá thực nghiệm toàn diện, so sánh kiến trúc mạng tích chập thời gian (TCN) với các kiến trúc mạng hồi quy (RNN) kinh điển như LSTM và GRU trên nhiều tác vụ mô hình hóa chuỗi, bao gồm các bài kiểm tra tổng hợp, mô hình hóa âm

nhạc và ngôn ngữ. Kết quả cho thấy TCN vượt trội hơn RNN về độ chính xác, khả năng ghi nhớ dài hạn, tính toán song song, tính linh hoạt và hiệu quả bộ nhớ, đồng thời có quá trình huấn luyện ổn định hơn. Nghiên cứu này đề xuất rằng mạng tích chập nên được xem là một lựa chọn khởi đầu tự nhiên cho các tác vụ mô hình hóa chuỗi, thay vì mặc định sử dụng RNN.

- Toán tử tích chập cho phép thực hiện 1 số sửa đổi đơn giản để tính ra các quan sát 1 phần, tương tự với cách thức được sử dụng trong xử lý ảnh.

Bài báo của Liu et al. (2018) [6] đã giới thiệu khái niệm tích chập một phần cho các tác vụ phục hồi hình ảnh. Tích chập một phần được thiết kế để xử lý các lỗ hổng không đều hoặc các vùng bị thiếu trong hình ảnh bằng cách điều chỉnh phép toán tích chập để chỉ xem xét các pixel hợp lệ trong trường tiếp nhận. Cách tiếp cận này cho phép tái tạo hiệu quả nội dung hình ảnh bị thiếu.

- Khác với tích chập 1 phần thông thường (vanilla partial convolution - PartialCNN), nhóm tác giả đề xuất sử dụng tích chập thời gian 1 phần (partial temporal convolution - PartialTCN) với mỗi chuỗi thời gian khác nhau, nghĩa là các tham số của PartialTCN có thể chia sẻ giữa các chuỗi thời gian khác nhau. Lợi ích ở đây sẽ có 2 điều:

- Chúng ta sẽ phân tách mô hình phụ thuộc thời gian ra khỏi mô hình tương quan không gian, điều này giúp cho việc khám phá các mẫu thời gian không phụ thuộc vào những biến cụ thể (invariants to instances) và đồng thời đẩy mạnh về mặt thống kê (enhancing the statistical strength)
- PartialTCN có thể chia sẻ giữa những biến và điều này giúp tiết kiệm tham số mô hình hơn (parameter-efficient model).

- Thêm vào đó, họ cũng chấp nhận multi-scale strategy để phát triển Multi-Scale Instance PartialTCN, có khả năng xử lý các mẫu dữ liệu thiếu từ các thang thời gian khác nhau một cách hiệu quả.

- Để khám phá sâu hơn về Multi-Scale Instance PartialTCN Module, chúng ta sẽ đi tìm hiểu về thành phần nhỏ hơn cấu tạo nên nó được gọi là Instance PartialTCN.

- Đầu tiên, giả sử chúng ta có 1 kích thước kernel là K , TCN áp dụng cùng 1 phép biến đổi tuyến tính giữa các cửa sổ thời gian khác nhau dựa trên giả định bất biến dịch chuyển thời gian (time translation invariance), như là:

$$x' = x_{t-K:t}W + b \quad (4.3)$$

Ở đây, ta có $x_{t-K:t} : t \in \mathbb{R}^{K \times D_i}$ là 1 cửa sổ thời gian với D_i chức năng đầu vào, $x' \in \mathbb{R}^{D_o}$ sẽ là bảng đặc trưng đầu ra (output feature map) tại vị trí $t - 1$, và

$W \in \mathbb{R}^{K \times D_i \times D_o}$ cũng như $b \in \mathbb{R}^{D_o}$ là những tham số tích chập.

- Lấy cảm hứng từ thành công của tích chập cục bộ trong những tác vụ thị giác, nhóm tác giả đã giới thiệu Instance PartialTCN đến mô hình phụ thuộc thời gian với chuỗi thời gian quan sát 1 phần để tính toán các giá trị bị thiếu như:

$$x' = \begin{cases} \frac{K}{\text{sum}(m_{t-K:t})}(x_{t-K:t} \odot m_{t-K:t})W + b & \text{nếu } \text{sum}(m_{t-K:t}) > 0 \\ 0 & \text{ngược lại,} \end{cases} \quad (4.4)$$

trong đó, \odot biểu thị tích Hadamard (Hadamard product).

- Instance PartialTCN chỉ tập trung các bước thời gian có quan sát để tính toán những bảng đặc trưng mới và nhân tố $\frac{K}{\text{sum}(m_{t-K:t})}$ hiệu chỉnh kết quả tính toán về cùng 1 độ lớn với các tích chập trên những quan sát hoàn chỉnh.
- Bằng cách này, những mẫu bị thiếu được tích hợp vào mô hình hóa động lực thời gian.
- Khi tích hợp thời gian diễn ra, những bước thời gian với những giá trị thiếu sẽ có cơ hội để nhận những thông tin tốt đến từ những vùng lân cận. Để tính toán điều này, mẫu dữ liệu thiếu \mathbf{m} được cập nhật lại như sau:

$$m_{t-1} = \begin{cases} 1 & \text{nếu } \text{sum}(m_{t-K:t}) > 0, \\ 0 & \text{ngược lại.} \end{cases} \quad (4.5)$$

Nói cách khác, bước thời gian thứ $t-1$ được coi là đủ nếu chúng ta thu thập giá trị từ tại cửa sổ thời gian hiện tại $[t-K, t)$. Mẫu giá trị m được coi là được lấp đầy dần dần theo quá trình tích chập.

- Đến với Multi-Scale Instance PartialTCN, để nắm được những phụ thuộc thời gian trong chuỗi thời gian, nhóm tác giả Xiaodan Chen đề xuất tích hợp chập đa tỉ lệ với những kích thước kernel khác nhau vào Instance PartialTCN.
- Cụ thể trong bài báo này, họ sẽ sử dụng kích thước $1 \times 3, 1 \times 5, 1 \times 7$. Kết quả là những kernel khác nhau sẽ tạo ra những mẫu thiếu cập nhật khác nhau $m_i^{(n)} \in \mathbb{R}^H$ cho mỗi phần tử n , $1 \leq i \leq N_{ker}$ và N_{ker} là số lượng kernel. Vì vậy, nhóm tác giả đã đề xuất cách tổng hợp những mẫu thiếu được tạo ra bởi những kernel khác nhau bằng phép lấy

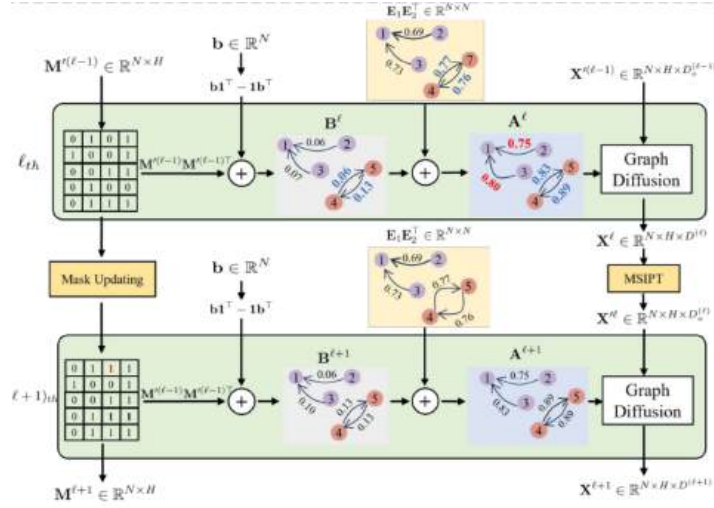
giá trị lớn nhất (max pooling) như sau:

$$m_i^{(n)} = \max(m_i^{(n)}), 1 \leq i \leq N_{ker} \quad (4.6)$$

- Mẫu thiếu tổng hợp $m^{(n)}$ (aggregated missing pattern) sẽ được sử dụng trong những mô đun tích hợp đồ thị (graph convolution module) tiếp theo để truyền thông tin dọc theo các chiều không gian. Bằng việc áp dụng Multi-Scale Instance PartialTCN cho mỗi biến $x^{(n)}$ ($n = 1, 2, \dots, N$), ta sẽ biến đổi bằng đặc trưng đầu vào $X^{(\ell-1)} \in \mathbb{R}^{N \times H \times D^{\ell-1}}$ và mẫu thiếu $M^{(\ell-1)} \in \mathbb{R}^{(N \times H)}$ thành $X'^{(\ell-1)} \in \mathbb{R}^{N \times H \times D_o^{(\ell-1)}}$ và mẫu thiếu tương ứng $M'^{(\ell-1)} \in \mathbb{R}^{N \times H}$.

4.2.3. Biased GCN Module

- Multi-Scale Instance PartialTCN tập trung vào việc bắt những biến thời gian (temporal dynamics) của từng phần tử riêng lẻ, bỏ qua mối tương quan không gian (spatial correlation) giữa các biến thể. Tuy nhiên, mục tiêu quan trọng nhất của chúng ta chính là mô hình hóa được sự tương quan không gian và phụ thuộc thời gian để dự đoán được chuỗi thời gian đa biến.
- Ở bài báo này, nhóm tác giả đã sử dụng mạng tích hợp đồ thị (graph convolution network) để khám phá những cấu trúc không gian của của bản đồ đặc trưng theo thời gian (temporally fused feature map) $X'^{(\ell-1)}$ và ma trận mặt nạ được cập nhật $M'^{(\ell-1)}$ được tạo ra bởi Multi-Scale Instance PartialTCN.
- Những mạng đồ thị nơ ron được khám phá để mô hình hóa tương quan không gian cho việc dự đoán chuỗi thời gian đa biến trong ngữ cảnh này được sử dụng bởi cấu trúc đồ thị xác định trước (predefined structures) và cấu trúc đồ thị học tự động (adaptively-learned graph structures). Trong đó, mỗi chuỗi thời gian được coi là 1 nút của đồ thị.
- Ngược lại với những giải pháp đã tồn tại, chúng ta sẽ cân nhắc đưa bias vào cấu trúc đồ thị để tính toán những giá trị bị thiếu, dẫn đến Biased GCN. Do đó, phương pháp này có thể chuyển biến những hiệu quả tốt trong ngữ cảnh dữ liệu bị thiếu. Biased GCN mô đun được trình bày ở Hình 4.
- Trong bài báo này, nhóm nghiên cứu đã chọn phương pháp học cấu trúc đồ thị thích nghi bởi vì nó linh hoạt hơn để ứng dụng vào những trường hợp khi mà cấu trúc đồ thị không có sẵn. Cụ thể, họ chọn cấu trúc đồ thị hoặc là ma trận kề A bằng việc sử dụng



Hình 4.4: Minh họa cho khối Biased GCN

2 ma trận nhưng có thể học được $E_1, E_2 \in \mathbb{R}^{N \times D_{\text{node}}}$ như sau:

$$A = \text{ReLU}(\tanh(E_1 E_2^T)) \quad (4.7)$$

Với hàng thứ i của E_1 (tương tự với E_2) có kí hiệu là $e_i^{(1)}$ (tương tự với $e_i^{(2)}$) là vector nhưng của chuỗi thời gian thứ i . Giá trị $e_i^{(1)\top} e_j^{(2)}$ đo lường mức tương quan từ node thứ i đến node thứ j .

- Lí do nhóm nghiên cứu chọn 2 ma trận nhưng này thay vì 1 ma trận E và 1 hàm tính toán $A = \text{ReLU}(\tanh(EE^T))$ chúng là những quan hệ tương quan này rất có thể không đối xứng trong thực tế. Nghĩa là mối quan hệ giữa 2 chuỗi thời gian có thể không giống nhau theo 2 hướng.
- Ma trận A sẽ được sử dụng bởi những phép tính chập đồ thị để tổng hợp thông tin và hỗ trợ cho nhiệm vụ dự báo cuối cùng. Do đó, 2 ma trận E_1 và E_2 có thể được học toàn bộ mô hình cùng 1 lúc.
- Tuy nhiên, công thức thứ 7 thì thất bại trong việc tính toán những mẫu dữ liệu thiếu. Cường độ lan truyền thông tin nên thay đổi theo các mẫu dữ liệu bị thiếu và nhóm tác giả chọn inner product để tính toán chúng:

$$A = \text{ReLU}(\tanh(E_1 E_2^T)) + \beta \text{softmax}(M_{t-H:t} : M_{t-H:t}^T) \quad (4.8)$$

trong đó, số hạng đầu tiên biểu hiện cường độ tương quan không gian toàn cục để đưa

ra cường độ truyền thông điệp toàn cục giữa các nút, số hạng thứ 2 thì đặc trưng cho 1 cửa sổ thời gian cụ thể $[t - H, t)$ có thể được nhìn nhận như là 1 độ lệch cụ thể cho cửa sổ thời gian đối với mẫu bị thiếu hiện tại trong quá trình khuếch tán đồ thị, β là một tham số toàn cục có thể kiểm soát được cường độ của sự chính xác.

- Sự lan truyền thông tin nên được định hướng và nhiều thông tin hơn nên đi từ các nút có ít giá trị bị thiếu đến nút với nhiều giá trị bị thiếu hơn, nhưng hạng tử thứ 2 là 1 ma trận đối xứng và không thể phản ánh được điều này.
- Để sửa chữa điều này, chúng ta có thể thêm mỗi nút một độ lệch vô hướng có thể học được b_i và sử dụng $b_i - b - j$ để điều chỉnh hướng tới các bất đối xứng. $b \in \mathbb{R}^2$ chính là một hạng tử độ lệch có thể học được, nhóm tác giả có đề xuất về việc học những cấu trúc đồ thị như sau:

$$\begin{aligned} B &= \text{softmax}(M_{t-H:t} M_{t-H:t}^\top + b \mathbf{1}^\top - \mathbf{1} b^\top) \\ A &= \text{ReLU}(\tanh(E_1 E_2^\top)) + \beta B \end{aligned} \quad (4.9)$$

Trong đó, số $\mathbf{1}$ là một vectơ toàn bộ một có độ dài N . Như được hiển thị trong Hình 4, độ mạnh truyền thông điệp toàn cục (global message passing strengths) giữa nút 4 và nút 5 được điều chỉnh bởi độ lệch cụ thể theo cửa sổ thời gian.

- Để đảm bảo tính thừa của cấu trúc, nhóm nghiên cứu giới hạn các phần tử nhỏ của A về 0 bằng cách chỉ giữ lại các hàng xóm của nút i với độ mạnh tương quan top- k và sử dụng A đã được giới hạn trong phép toán tích chập đồ thị để tổng hợp thông tin.
- Tương tự như Instance PartialTCN, nhóm tác giả đã đề xuất cập nhật các mẫu còn thiếu của nút i sau khi tổng hợp thông tin từ các hàng xóm không gian của nó như sau:

$$m^{(i)} = \max(m^{(j)}), j \in i \cup N_i \quad (4.10)$$

trong đó N_i biểu thị các hàng xóm của nút i trong đồ thị. Quá trình cập nhật mẫu còn thiếu được minh họa với nút 1 trong Hình 4.4.

- Bây giờ, ta hãy cùng xem xét khối thứ ℓ của mô hình, nó thực hiện tích chập đồ thị để lan truyền thông tin như sau.

$$\mathbf{X}^{(\ell)} = (\mathbf{I} + \mathbf{D}_o^{-1} \mathbf{A} + \mathbf{D}_i^{-1} \mathbf{A}^\top) \mathbf{X}'^{(\ell-1)} \boldsymbol{\Theta}^{(\ell)} + \mathbf{b}^{(\ell)} \quad (4.11)$$

trong đó $\mathbf{X}'^{(\ell-1)}$ là đầu ra của Multi-Scale Instance PartialTCN trong khối thứ ℓ , \mathbf{D}_i

và \mathbf{D}_o lần lượt là ma trận bậc vào và bậc ra của \mathbf{A} , và $\Theta^{(\ell)}$ và $\mathbf{b}^{(\ell)}$ là các tham số tích chập đồ thị của khối thứ ℓ . $\mathbf{X}^{(\ell)}$ và $\mathbf{M}^{(\ell)}$ sau đó sẽ được đưa vào khối tiếp theo làm đầu vào.

4.2.4. Kiến trúc phân cấp - Hierarchical Architecture

- Bằng cách xếp chồng L lớp Biased TCGBlock, chúng ta có thể tăng cường cả trường tiếp nhận không gian và thời gian (spatial and temporal receptive fields) của mô hình.
- Chúng ta khởi tạo $\mathbf{X}^{(0)}$ và $\mathbf{M}^{(0)}$ như là quan sát một phần ban đầu $\mathbf{X} \in \mathbb{R}^{N \times H \times D}$ và mẫu còn thiếu tương ứng của nó $\mathbf{M} \in \mathbb{R}^{N \times H}$, và đầu ra của khối thứ L là $\mathbf{X}^{(L)} \in \mathbb{R}^{N \times H \times D^{(L)}}$ và $\mathbf{M}^{(L)} \in \mathbb{R}^{N \times H}$.
- $\mathbf{X}^{(L)}$ kết hợp cả các đặc trưng không gian và thời gian, sẽ được sử dụng để tạo ra dự đoán nhiều bước \hat{Y} đồng thời bằng một phép biến đổi tuyến tính.
- Tại lớp mặt nạ thứ ℓ , $\mathbf{M}^{(\ell)}$ được cập nhật dần dần khi thông tin chảy từ dưới lên trên, và các tham số mô hình được học bằng cách tối ưu hóa hàm mất mát dự đoán trong công thức số 2.

4.3. Nghiên cứu cắt bỏ

- Nhóm tác giả đã tiến hành các nghiên cứu cắt bỏ để đánh giá hiệu quả của các mô-đun được đề xuất, cụ thể là mô-đun Multi-Scale Instance PartialTCN (MSIPT) và mô-đun Biased GCN (BGCN). Kết quả được thể hiện trong Hình 4.5. Họ chia mô-đun MSIPT hoặc BGCN thành hai quy trình riêng biệt. Phần đầu tiên (Phương trình 4.6 hoặc Phương trình 4.10) liên quan đến quá trình cập nhật mặt nạ (MUP), trong khi phần thứ hai (Phương trình 4.4 hoặc Phương trình 4.9) liên quan đến quá trình tổng hợp thông tin (IAP).
- Đầu tiên, họ thực hiện các nghiên cứu cắt bỏ (không có MSIPT, không có BGCN và BiTGraph) để đánh giá tầm quan trọng chung của UID và MUP trên các chiều không gian và thời gian. Kết quả cho thấy khi họ áp dụng MSIPT hoặc BGCN, hiệu suất giảm đáng kể, điều này có thể được giải thích là do MUP xây dựng một đường dẫn truyền thông tin hoàn chỉnh giữa các chiều không gian và thời gian và việc thiếu bất kỳ mô-đun nào sẽ cắt đứt luồng thông tin giữa hai chiều.

- Để xác nhận thêm hiệu quả của IAP và MUP, nhóm nghiên cứu đã tiến hành các nghiên cứu cắt bỏ bằng cách sửa đổi IAP. Đầu tiên, họ thay thế phép tích chập thời gian trong mô-đun MSIPT bằng phép tích chập tiêu chuẩn (ví dụ: Phương trình 4.3). Tiếp theo, nhóm tác giả thay đổi việc tạo ma trận liên kề A bằng cách sử dụng Phương trình 4.7. Kết quả được hiển thị trong hàng thứ tư và thứ năm của bảng. Như chúng ta có thể thấy, việc thay thế các hoạt động không gian hoặc thời gian trong IAP dẫn đến giảm hiệu suất đáng kể, điều này tiếp tục xác minh tính hiệu quả của các mô-đun được đề xuất của chúng ta trong việc xử lý các giá trị bị thiếu.

Missing Rate	Model	Metr		Electricity		PEMS	
		MAE	RMSE	MAE	RMSE	MAE	RMSE
0.2	TCGNet	6.25 \pm 0.01	11.01 \pm 0.02	255.46 \pm 2.86	2026.43 \pm 35.79	1.94 \pm 0.00	4.35 \pm 0.02
	w/o. MSIPT	6.34 \pm 0.04	11.25 \pm 0.05	279.73 \pm 4.91	2159.16 \pm 28.74	1.97 \pm 0.02	4.40 \pm 0.04
	w/o. BGCN	6.26 \pm 0.03	11.33 \pm 0.03	263.69 \pm 3.32	2029.16 \pm 25.56	1.94 \pm 0.00	4.37 \pm 0.03
	w/o. Eq. 9	6.14 \pm 0.02	10.72 \pm 0.02	250.75 \pm 5.01	2020.13 \pm 36.75	1.94 \pm 0.02	4.33 \pm 0.03
	w/o. Eq. 4	6.12 \pm 0.01	10.91 \pm 0.03	246.18 \pm 2.89	2003.22 \pm 35.09	1.93 \pm 0.01	4.30 \pm 0.02
	BiTGraph	6.04 \pm 0.02	10.69 \pm 0.02	243.23 \pm 2.12	1834.18 \pm 15.36	1.90 \pm 0.01	4.28 \pm 0.01
0.4	TCGNet	6.41 \pm 0.03	11.14 \pm 0.04	284.39 \pm 5.26	2323.03 \pm 40.15	1.99 \pm 0.01	4.47 \pm 0.03
	w/o. MSIPT	6.48 \pm 0.02	11.20 \pm 0.03	299.34 \pm 4.17	2361.79 \pm 37.82	2.02 \pm 0.02	4.50 \pm 0.03
	w/o. BGCN	6.40 \pm 0.02	11.20 \pm 0.02	291.81 \pm 3.87	2337.69 \pm 31.98	1.98 \pm 0.01	4.45 \pm 0.02
	w/o. Eq. 9	6.18 \pm 0.01	10.81 \pm 0.04	282.12 \pm 2.88	2236.82 \pm 30.26	1.98 \pm 0.01	4.32 \pm 0.02
	w/o. Eq. 4	6.25 \pm 0.00	10.87 \pm 0.02	280.30 \pm 2.73	2277.50 \pm 28.49	1.97 \pm 0.01	4.34 \pm 0.01
	BiTGraph	6.13 \pm 0.01	10.76 \pm 0.02	270.14 \pm 3.77	2091.88 \pm 30.49	1.96 \pm 0.00	4.34 \pm 0.02
0.6	TCGNet	6.48 \pm 0.02	11.10 \pm 0.05	313.60 \pm 3.29	2372.36 \pm 36.19	2.04 \pm 0.02	4.55 \pm 0.02
	w/o. MSIPT	6.65 \pm 0.03	11.50 \pm 0.04	332.39 \pm 3.82	2469.15 \pm 33.63	2.09 \pm 0.02	4.62 \pm 0.04
	w/o. BGCN	6.65 \pm 0.02	11.94 \pm 0.01	322.68 \pm 2.74	2487.22 \pm 25.39	2.03 \pm 0.00	4.49 \pm 0.01
	w/o. Eq. 9	6.35 \pm 0.03	11.06 \pm 0.02	308.59 \pm 3.97	2366.39 \pm 32.16	2.03 \pm 0.01	4.52 \pm 0.02
	w/o. Eq. 4	6.38 \pm 0.02	10.84 \pm 0.02	301.25 \pm 2.05	2312.39 \pm 22.46	2.03 \pm 0.02	4.54 \pm 0.00
	BiTGraph	6.32 \pm 0.01	10.93 \pm 0.03	295.23 \pm 2.75	2239.06 \pm 26.39	1.99 \pm 0.01	4.47 \pm 0.01

Hình 4.5: Kết quả thử nghiệm của BiTGraph đối với từng trường hợp đã đề cập.

Kết quả cuối cùng

5.1. Cơ sở so sánh

- Nhóm tác giả có so sánh BiTGraph với các phương pháp dự báo tiên tiến nhất cũng như một số phương pháp cổ điển. BRITS (Cao et al., 2018)[3], SPIN (Marisca et al., 2022)[7], GRIN (Cini et al., 2022)[4], GCN-M (Zuo et al., 2023)[13], CRUs (Schirmer et al., 2022)[8] là các phương pháp dự báo đại diện được thiết kế đặc biệt cho chuỗi thời gian có giá trị bị thiếu.
- Đồng thời, họ cũng bao gồm ba phương pháp dựa trên Transformer, vanilla Transformer (Zerveas et al., 2021)[11], STWA (Cirstea et al., 2022)[5] và FEDformer (Zhou et al., 2022)[12], cũng như hai phương pháp dựa trên Spatial-Temporal GNNs, AGCRN (Bai et al., 2020)[1] và MTGNN (Wu et al., 2020)[10]. Vì năm phương pháp này yêu cầu đầu vào hoàn chỉnh để thực hiện dự đoán, họ nghiên cứu hai biến thể của chúng, cụ thể là điền các mục bị thiếu bằng số không và các giá trị được gán bởi TimesNet (Wu et al., 2023)[9], phương pháp gán chuỗi thời gian tiên tiến nhất. Nhóm nghiên cứu cũng biểu thị các biến thể tương ứng là Model_0 và Model_t . Các mặt nạ bị thiếu được đưa vào dưới dạng hiệp biến để hướng dẫn dự báo cho năm phương pháp cơ sở sau.
- Đối với BRITS, SPIN, GRIN, CRUs, AGCRN, MTGNN, FEDformer và STWA, họ sử dụng mã nguồn do tác giả của các mô hình đã phát hành. Trong khi đối với Transformer, nhóm tác giả sử dụng phiên bản được triển khai trong FEDformer. Đây là thông tin chi tiết của các phương pháp đã đề cập trước đó:
 - BRITS: Đây là một mô hình gán chuỗi thời gian kết hợp Mạng thần kinh tái phát hai chiều với cơ chế phân rã thời gian để thiết lập mối quan hệ giữa các giá trị bị thiếu và dữ liệu quan sát được.
 - SPIN: Xử lý các giá trị bị thiếu bằng cách xây dựng một đồ thị được kết nối thừa thớt trong cả chiều không gian và thời gian.
 - GRIN: Kết hợp mạng thần kinh tái phát (RNN) và mạng thần kinh đồ thị (GNN) để nắm bắt các phụ thuộc giữa các chuỗi và trong chuỗi để xây dựng mối quan hệ

- giữa các giá trị bị thiếu và các giá trị quan sát được.
- GCN-M: Xem xét các đặc trưng không gian-thời gian cục bộ và các mẫu lịch sử toàn cục trong một mạng bộ nhớ dựa trên sự chú ý.
 - CRUs: Kết hợp bộ lọc Kalman và khung mã hóa-giải mã để cập nhật các trạng thái ẩn liên tục.
 - AGCRN: Đồ thị thích ứng và các mẫu cụ thể theo nút được học tương ứng bởi các nhúng nút và phân tích ma trận.
 - MTGNN: Phương pháp này xây dựng một ma trận tương quan không gian đối xứng lệch và sử dụng Mạng Convolutional Thời gian (TCN) và GCN để nắm bắt các phụ thuộc trong và giữa các chuỗi.
 - Transformer: Mô hình tuần tự cổ điển sử dụng một chồng các khối tự chú ý để nắm bắt các phụ thuộc thời gian trong chuỗi thời gian.
 - FEDformer: Phương pháp này phân tách chuỗi thời gian thành các thành phần theo mùa và xu hướng và sử dụng cơ chế tự chú ý trong miền tần số.
 - STWA: Xây dựng các nhúng nhận biết không gian-thời gian trong cơ chế tự chú ý và giới thiệu sự chú ý cửa sổ để giảm độ phức tạp.

5.2. Thiết lập đầu vào

Số lượng khối L của BiTGraph được đặt thành 3, số lượng hàng xóm gần nhất top-k được đặt thành 10 trong tất cả các thí nghiệm của nhóm nghiên cứu. Kích thước lô (batch size) là 32, tốc độ học (learning rate) là 0,001. Họ chia các bộ dữ liệu thành các bộ dữ liệu huấn luyện, xác thực và kiểm tra theo tỷ lệ 0,6/0,2/0,2 theo thứ tự thời gian. Kích thước cửa sổ tương lai F được đặt thành 24 cho tất cả các phương pháp và kích thước cửa sổ lịch sử H cho phương pháp đề xuất của họ là 24. Nhóm tác giả chọn kích thước cửa sổ lịch sử tốt nhất từ tập hợp 24, 48, 96 cho các phương pháp cơ sở và báo cáo kết quả tốt nhất của chúng. Tất cả các phương pháp được đào tạo trên GPU Nvidia V100. Phương pháp của họ được triển khai bằng PyTorch 2.0 và có sử dụng mã nguồn do các tác giả phát hành cho tất cả các phương pháp cơ sở. Nhóm nghiên cứu có điều chỉnh các siêu tham số của các phương pháp cơ sở để có được hiệu suất tốt nhất trên mỗi bộ dữ liệu và đánh giá hiệu suất của các phương pháp khác nhau về Sai số tuyệt đối trung bình (MAE), Sai số bình phương trung bình căn bậc hai (RMSE) và Sai số phần trăm tuyệt đối trung bình (MAPE).// Ngoài ra, để chạy thực nghiệm thử mô hình, hãy làm theo các bước sau:

- Bước 1: Sử dụng kaggle, google colab hoặc bất cứ môi trường nào có sử dụng được GPU đến từ NVIDIA.
- Bước 2: Nhập lệnh '!git clone https://github.com/DBLue04/BiTGraph-main.git' để clone repository về workplace của mình, sau đó chạy lệnh '%cd BiTGraph-main' để di chuyển vào thư mục BiTGraph-main. Lưu ý: Đây chỉ là github clone và tinh chỉnh lại để có hoạt động được, và chỉ chạy được trên 3 bộ dữ liệu chính là: Metr-LA, ETTH01 và PEMS.
- Bước 3: Đảm bảo đã cài đặt và cập nhật các thư viện yêu cầu trong 'requirement.txt'
- Bước 4: Chạy lệnh '!python main.py' để khởi chạy mô hình, sau đó thì chạy lệnh '!python test_forecasting.py' để lấy kết quả và đánh giá hiệu năng của mô hình.

5.3. Kết quả tổng quan

Method (0.1)	PEMS		ETTh1		BeijingAir	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
BRITS	3.06 ± 0.01	6.39 ± 0.02	1.76 ± 0.03	3.17 ± 0.03	45.79 ± 0.24	67.83 ± 0.21
SPIN	2.03 ± 0.00	4.62 ± 0.00	—	—	44.93 ± 0.03	68.13 ± 0.75
GRIN	2.63 ± 0.01	6.03 ± 0.06	—	—	45.96 ± 0.25	67.06 ± 1.30
GCNM	2.13 ± 0.02	5.29 ± 0.07	—	—	47.68 ± 0.11	68.29 ± 0.09
CRUs	3.21 ± 0.02	6.03 ± 0.01	2.80 ± 0.03	4.82 ± 0.05	56.92 ± 0.72	76.13 ± 0.87
AGCRN ₀	5.10 ± 0.07	10.07 ± 0.06	2.39 ± 0.04	4.76 ± 0.08	55.50 ± 0.12	81.31 ± 0.19
Transformer ₀	2.75 ± 0.07	6.15 ± 0.02	1.88 ± 0.05	3.25 ± 0.07	48.58 ± 0.07	69.50 ± 0.21
FEDformer ₀	2.61 ± 0.05	5.76 ± 0.10	1.69 ± 0.02	3.22 ± 0.04	49.65 ± 0.09	72.77 ± 0.13
STWA ₀	2.01 ± 0.04	4.57 ± 0.03	1.75 ± 0.00	3.15 ± 0.00	46.71 ± 0.14	70.10 ± 0.07
MTGNN ₀	2.02 ± 0.01	4.52 ± 0.04	1.58 ± 0.01	2.99 ± 0.02	44.37 ± 0.05	65.92 ± 0.06
AGCRN _t	5.08 ± 0.02	10.05 ± 0.00	2.16 ± 0.03	4.29 ± 0.05	47.08 ± 0.26	69.62 ± 0.37
Transformer _t	2.54 ± 0.02	6.05 ± 0.03	1.72 ± 0.01	3.25 ± 0.13	47.43 ± 0.16	69.69 ± 0.08
FEDformer _t	2.45 ± 0.03	5.43 ± 0.02	1.67 ± 0.02	3.22 ± 0.04	44.87 ± 0.16	66.54 ± 0.07
STWA _t	1.98 ± 0.02	3.51 ± 0.04	1.64 ± 0.01	3.04 ± 0.02	45.28 ± 0.13	68.93 ± 0.04
MTGNN _t	1.93 ± 0.01	3.35 ± 0.03	1.54 ± 0.01	2.96 ± 0.02	43.32 ± 0.02	65.81 ± 0.07
BiTGraph	1.56 ± 0.02	2.97 ± 0.02	1.51 ± 0.01	2.92 ± 0.02	42.11 ± 0.11	65.53 ± 0.23

Hình 5.1: Kết quả đầu ra của 3 bộ dữ liệu với độ thiếu dữ liệu là 0.1

5.4. Nhận xét

Dựa trên 3 bộ dữ liệu được chạy thực nghiệm của chính nhóm bao gồm: Metr-LA, ETTH01 và PEMS. Thì nhóm có những nhận xét như sau:

- Kết quả đầu ra của BiTGraph đa phần tốt hơn so với hầu hết cách phân kết quả khác.
- Tuy nhiên, sai số được đến từ nhóm tác giả có vẻ như chưa được chính xác lắm.
- Trung bình 1 lần chạy thực nghiệm sẽ kéo dài từ 20 phút đến 120 phút (Dựa vào kích thước bộ dữ liệu).

Method (0.4)	PEMS		ETTh1		BeijingAir	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
BRITS	3.08 ± 0.01	6.31 ± 0.03	1.82 ± 0.07	3.26 ± 0.06	46.20 ± 0.35	67.95 ± 0.06
SPIN	2.13 ± 0.10	4.81 ± 0.24	—	—	45.37 ± 0.55	67.38 ± 0.77
GRIN	2.81 ± 0.03	6.76 ± 0.14	—	—	46.79 ± 0.25	68.22 ± 0.14
GCN-M	2.26 ± 0.02	5.51 ± 0.03	—	—	50.61 ± 0.04	71.61 ± 0.17
CRUs	2.85 ± 0.02	5.65 ± 0.04	2.94 ± 0.05	5.01 ± 0.04	55.80 ± 0.46	76.41 ± 1.16
AGCRN ₀	5.09 ± 0.01	10.05 ± 0.02	3.19 ± 0.03	5.06 ± 0.07	55.52 ± 0.07	81.27 ± 0.03
Transformer ₀	2.86 ± 0.02	6.38 ± 0.07	2.36 ± 0.04	4.79 ± 0.08	53.11 ± 0.06	77.03 ± 0.17
FEDformer ₀	2.81 ± 0.05	5.81 ± 0.09	1.96 ± 0.02	3.61 ± 0.03	52.18 ± 0.05	74.78 ± 0.16
STWA ₀	2.12 ± 0.02	4.67 ± 0.03	1.90 ± 0.03	3.40 ± 0.03	47.38 ± 0.07	71.16 ± 0.06
MTGNN ₀	2.12 ± 0.01	4.50 ± 0.02	1.92 ± 0.02	3.43 ± 0.01	45.17 ± 0.08	66.73 ± 0.02
AGCRN _t	5.10 ± 0.00	10.06 ± 0.01	2.87 ± 0.04	5.24 ± 0.03	55.56 ± 0.07	82.67 ± 0.10
Transformer _t	2.80 ± 0.03	6.09 ± 0.07	1.88 ± 0.02	3.68 ± 0.08	48.28 ± 0.16	69.71 ± 0.09
FEDformer _t	2.74 ± 0.02	5.78 ± 0.03	1.86 ± 0.02	3.31 ± 0.01	46.96 ± 0.13	68.17 ± 0.08
STWA _t	2.07 ± 0.03	4.80 ± 0.02	1.81 ± 0.04	3.26 ± 0.07	45.69 ± 0.13	70.56 ± 0.19
MTGNN _t	2.05 ± 0.02	4.61 ± 0.07	1.67 ± 0.01	3.12 ± 0.01	44.29 ± 0.02	66.39 ± 0.08
BITGraph	1.96 ± 0.00	4.34 ± 0.02	1.64 ± 0.02	3.07 ± 0.02	43.13 ± 0.22	65.55 ± 0.24
Method (0.6)	PEMS		ETTh1		BeijingAir	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
BRITS	3.14 ± 0.03	6.22 ± 0.01	1.87 ± 0.11	3.24 ± 0.02	46.71 ± 0.17	67.93 ± 0.04
SPIN	2.18 ± 0.05	4.91 ± 0.15	—	—	44.28 ± 0.45	65.68 ± 0.08
GRIN	2.84 ± 0.01	6.61 ± 0.1	—	—	49.10 ± 1.01	71.43 ± 1.63
GCN-M	2.39 ± 0.02	5.32 ± 0.03	—	—	51.47 ± 0.05	77.41 ± 0.22
CRUs	2.97 ± 0.31	5.79 ± 0.09	3.05 ± 0.05	5.13 ± 0.04	56.63 ± 0.11	75.42 ± 1.76
AGCRN ₀	5.09 ± 0.01	10.06 ± 0.01	3.43 ± 0.02	5.81 ± 0.05	55.67 ± 0.06	82.10 ± 0.05
Transformer ₀	2.91 ± 0.03	6.31 ± 0.04	2.45 ± 0.03	4.30 ± 0.03	50.79 ± 0.11	71.73 ± 0.09
FEDformer ₀	2.85 ± 0.04	6.02 ± 0.06	2.06 ± 0.01	3.96 ± 0.02	55.32 ± 0.13	79.83 ± 0.03
STWA ₀	2.15 ± 0.02	4.64 ± 0.06	2.16 ± 0.02	4.86 ± 0.09	48.22 ± 0.07	70.09 ± 0.15
MTGNN ₀	2.13 ± 0.01	4.62 ± 0.02	2.36 ± 0.02	4.14 ± 0.03	47.05 ± 0.11	67.80 ± 0.06
AGCRN _t	5.10 ± 0.00	10.06 ± 0.01	3.35 ± 0.02	5.29 ± 0.04	55.45 ± 0.17	80.76 ± 0.04
Transformer _t	2.83 ± 0.04	5.22 ± 0.07	2.19 ± 0.01	4.29 ± 0.03	49.23 ± 0.04	72.38 ± 0.12
FEDformer _t	2.79 ± 0.02	6.31 ± 0.05	2.07 ± 0.01	4.26 ± 0.03	49.21 ± 0.06	68.27 ± 0.07
STWA _t	2.14 ± 0.02	4.67 ± 0.03	1.85 ± 0.02	3.37 ± 0.03	46.06 ± 0.03	69.92 ± 0.07
MTGNN _t	2.11 ± 0.01	4.59 ± 0.03	1.79 ± 0.02	3.29 ± 0.03	44.85 ± 0.02	67.78 ± 0.05
BITGraph	1.99 ± 0.01	4.47 ± 0.01	1.74 ± 0.00	3.21 ± 0.01	44.23 ± 0.15	64.20 ± 0.32
Method (0.8)	PEMS		ETTh1		BeijingAir	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
BRITS	3.26 ± 0.10	7.02 ± 0.03	2.14 ± 0.01	3.81 ± 0.09	46.75 ± 0.59	68.17 ± 0.87
SPIN	2.26 ± 0.01	5.03 ± 0.02	—	—	44.94 ± 0.13	66.58 ± 0.25
GRIN	2.96 ± 0.02	6.70 ± 0.12	—	—	52.07 ± 0.92	74.60 ± 1.85
GCN-M	2.54 ± 0.02	5.77 ± 0.04	—	—	52.57 ± 0.08	73.71 ± 0.14
CRUs	3.15 ± 0.01	6.07 ± 0.01	3.15 ± 0.11	5.31 ± 0.09	57.10 ± 1.45	73.18 ± 0.14
AGCRN ₀	5.09 ± 0.00	10.07 ± 0.02	3.18 ± 0.12	6.02 ± 0.02	55.59 ± 0.03	81.58 ± 0.10
Transformer ₀	2.97 ± 0.06	6.73 ± 0.08	2.81 ± 0.07	5.20 ± 0.06	52.47 ± 0.03	73.10 ± 0.16
FEDformer ₀	2.89 ± 0.04	6.37 ± 0.06	2.59 ± 0.03	5.78 ± 0.02	56.94 ± 0.04	81.88 ± 0.09
STWA ₀	2.34 ± 0.04	5.16 ± 0.02	2.56 ± 0.02	4.41 ± 0.03	52.07 ± 0.11	74.95 ± 0.08
MTGNN ₀	2.39 ± 0.02	5.18 ± 0.09	2.46 ± 0.04	5.30 ± 0.02	53.05 ± 0.02	73.57 ± 0.10
AGCRN _t	5.10 ± 0.01	10.07 ± 0.01	3.16 ± 0.09	5.25 ± 0.12	55.48 ± 0.05	78.86 ± 0.13
Transformer _t	2.85 ± 0.07	6.29 ± 0.11	2.51 ± 0.01	4.98 ± 0.07	50.09 ± 0.06	72.00 ± 0.07
FEDformer _t	2.76 ± 0.03	6.18 ± 0.02	2.55 ± 0.02	4.69 ± 0.06	49.83 ± 0.15	71.22 ± 0.08
STWA _t	2.27 ± 0.02	5.00 ± 0.01	2.11 ± 0.02	3.73 ± 0.04	46.97 ± 0.06	72.93 ± 0.07
MTGNN _t	2.21 ± 0.01	4.89 ± 0.01	2.01 ± 0.01	3.73 ± 0.02	45.93 ± 0.04	68.17 ± 0.11
BITGraph	2.15 ± 0.01	4.73 ± 0.02	1.91 ± 0.01	3.54 ± 0.01	45.47 ± 0.17	66.98 ± 0.29

Hình 5.2: Kết quả đầu ra của 3 bộ dữ liệu với độ thiếu dữ liệu là 0.4, 0.6 và 0.8

Method ($r = 0.1$)	MAE	Metr-LA RMSE	MAPE	MAE	Electricity RMSE	MAPE
BRITS	8.21 \pm 0.01	12.63 \pm 0.10	18.04 \pm 0.20	1027.47 \pm 3.92	10157.39 \pm 64.012	47.54 \pm 0.01
SPIN	6.36 \pm 0.00	11.07 \pm 0.01	12.00 \pm 0.00	—	—	—
GRIN	6.69 \pm 0.02	12.27 \pm 0.00	15.88 \pm 0.01	—	—	—
GCN-M	6.72 \pm 0.01	12.33 \pm 0.01	13.06 \pm 0.02	—	—	—
CRU	10.51 \pm 0.03	13.00 \pm 0.63	19.71 \pm 0.31	334.40 \pm 31.80	2923.44 \pm 39.73	24.99 \pm 0.72
AGCRN ₀	14.66 \pm 0.01	14.00 \pm 0.02	29.30 \pm 0.13	1361.11 \pm 8.39	12569.27 \pm 30.09	62.54 \pm 0.31
Transformer ₀	7.07 \pm 0.05	12.97 \pm 0.08	16.78 \pm 0.13	289.65 \pm 3.39	2296.17 \pm 27.77	25.08 \pm 0.32
FEDformer ₀	6.96 \pm 0.03	12.37 \pm 0.08	16.22 \pm 0.14	337.16 \pm 5.25	2713.72 \pm 33.37	30.24 \pm 0.28
STWA ₀	6.22 \pm 0.04	14.64 \pm 0.06	12.71 \pm 0.05	269.81 \pm 5.95	2039.64 \pm 24.06	22.39 \pm 0.33
MTGNN ₀	6.25 \pm 0.06	10.68 \pm 0.07	12.18 \pm 0.05	256.98 \pm 5.12	1974.50 \pm 14.69	20.95 \pm 0.26
AGCRN _t	13.72 \pm 0.06	13.11 \pm 0.23	27.06 \pm 0.18	1109.57 \pm 3.95	10794.08 \pm 34.26	57.95 \pm 0.18
Transformer _t	6.90 \pm 0.07	12.81 \pm 0.04	16.49 \pm 0.06	265.76 \pm 6.03	2064.82 \pm 23.51	20.06 \pm 0.27
FEDformer _t	6.61 \pm 0.03	11.09 \pm 0.10	13.23 \pm 0.06	283.63 \pm 2.15	2269.11 \pm 19.86	23.23 \pm 0.18
STWA _t	6.17 \pm 0.02	10.82 \pm 0.09	12.14 \pm 0.04	248.87 \pm 3.79	1945.16 \pm 23.20	18.20 \pm 0.16
MTGNN _t	6.10 \pm 0.02	10.69 \pm 0.03	12.02 \pm 0.08	254.67 \pm 3.22	1994.07 \pm 25.99	22.41 \pm 0.26
BiTGraph	5.96 \pm 0.01	10.71 \pm 0.00	11.13 \pm 0.02	231.70 \pm 1.76	1823.18 \pm 25.49	17.93 \pm 0.27
Method ($r = 0.2$)	MAE	Metr-LA RMSE	MAPE	MAE	Electricity RMSE	MAPE
BRITS	8.32 \pm 0.02	13.18 \pm 0.10	18.26 \pm 0.71	1029.30 \pm 1.10	10126.175 \pm 30.57	47.73 \pm 0.35
SPIN	6.46 \pm 0.07	11.21 \pm 0.05	12.98 \pm 0.03	—	—	—
GRIN	6.80 \pm 0.02	12.24 \pm 0.12	16.18 \pm 0.24	—	—	—
GCN-M	6.78 \pm 0.03	11.12 \pm 0.04	13.50 \pm 0.02	—	—	—
CRUs	10.80 \pm 0.02	12.49 \pm 0.15	19.66 \pm 0.54	464.66 \pm 4.14	5276.49 \pm 53.36	25.64 \pm 0.53
AGCRN ₀	14.88 \pm 0.05	14.21 \pm 0.04	28.94 \pm 0.07	1307.62 \pm 4.53	13217.78 \pm 26.81	62.65 \pm 0.22
Transformer ₀	7.14 \pm 0.06	13.08 \pm 0.08	17.07 \pm 0.09	296.03 \pm 5.77	2432.09 \pm 22.15	29.14 \pm 0.27
FEDformer ₀	7.09 \pm 0.03	12.75 \pm 0.14	16.73 \pm 0.19	368.29 \pm 3.71	2574.37 \pm 25.89	31.29 \pm 0.31
STWA ₀	6.24 \pm 0.07	10.99 \pm 0.11	12.89 \pm 0.13	272.60 \pm 7.35	2263.55 \pm 24.10	28.52 \pm 0.26
MTGNN ₀	6.34 \pm 0.07	10.96 \pm 0.10	12.51 \pm 0.19	274.68 \pm 5.56	2016.44 \pm 13.77	28.54 \pm 0.19
AGCRN _t	13.72 \pm 0.06	13.11 \pm 0.23	27.06 \pm 0.18	1049.23 \pm 12.06	11751.49 \pm 20.67	57.76 \pm 0.18
Transformer _t	6.90 \pm 0.08	12.98 \pm 0.13	16.49 \pm 0.21	280.12 \pm 6.78	2274.28 \pm 25.18	28.74 \pm 0.35
FEDformer _t	6.89 \pm 0.06	11.75 \pm 0.17	16.01 \pm 0.09	313.59 \pm 4.96	2666.93 \pm 26.31	32.83 \pm 0.23
STWA _t	6.20 \pm 0.02	10.71 \pm 0.11	12.26 \pm 0.18	261.92 \pm 4.65	2089.65 \pm 19.35	27.37 \pm 0.26
MTGNN _t	6.13 \pm 0.02	10.76 \pm 0.07	12.11 \pm 0.19	269.25 \pm 5.27	2175.24 \pm 12.49	27.71 \pm 0.78
BiTGraph	6.04 \pm 0.02	10.69 \pm 0.02	11.69 \pm 0.11	243.23 \pm 2.12	1834.18 \pm 15.36	27.38 \pm 0.46
Method ($r = 0.4$)	MAE	Metr-LA RMSE	MAPE	MAE	Electricity RMSE	MAPE
BRITS	8.38 \pm 0.08	12.97 \pm 0.11	18.39 \pm 0.28	1029.73 \pm 1.48	10136.39 \pm 63.63	47.96 \pm 0.56
SPIN	6.52 \pm 0.07	11.94 \pm 0.41	13.22 \pm 1.00	—	—	—
GRIN	6.91 \pm 0.09	12.60 \pm 0.21	16.59 \pm 0.18	—	—	—
GCN-M	7.09 \pm 0.01	12.42 \pm 0.03	17.06 \pm 0.04	—	—	—
CRUs	10.94 \pm 0.08	13.18 \pm 0.44	20.13 \pm 0.23	496.95 \pm 6.03	5397.31 \pm 52.52	27.94 \pm 0.33
AGCRN ₀	14.87 \pm 0.04	14.30 \pm 0.09	29.92 \pm 0.06	1526.90 \pm 13.77	14823.39 \pm 21.68	68.73 \pm 0.41
Transformer ₀	7.25 \pm 0.04	12.97 \pm 0.06	17.72 \pm 0.08	310.88 \pm 4.67	2586.69 \pm 22.73	31.79 \pm 0.15
FEDformer ₀	7.15 \pm 0.02	12.89 \pm 0.07	16.91 \pm 0.12	406.17 \pm 8.91	3606.49 \pm 27.73	33.14 \pm 0.33
STWA ₀	6.37 \pm 0.05	11.19 \pm 0.06	13.13 \pm 0.16	292.47 \pm 4.64	2764.34 \pm 20.06	29.07 \pm 0.16
MTGNN ₀	6.34 \pm 0.05	11.10 \pm 0.03	12.79 \pm 0.08	305.46 \pm 6.77	2576.44 \pm 25.51	23.15 \pm 0.37
AGCRN _t	12.73 \pm 0.02	12.49 \pm 0.14	24.13 \pm 0.16	1283.27 \pm 8.49	13743.42 \pm 49.38	58.62 \pm 0.36
Transformer _t	6.99 \pm 0.06	12.49 \pm 0.13	16.45 \pm 0.08	300.43 \pm 10.17	2529.26 \pm 19.14	28.86 \pm 0.20
FEDformer _t	7.10 \pm 0.05	12.63 \pm 0.13	16.62 \pm 0.06	330.90 \pm 7.76	2711.30 \pm 22.31	29.24 \pm 0.18
STWA _t	6.28 \pm 0.03	10.93 \pm 0.14	12.68 \pm 0.07	289.59 \pm 6.13	2355.34 \pm 17.67	28.29 \pm 0.31
MTGNN _t	6.26 \pm 0.05	10.90 \pm 0.10	12.49 \pm 0.04	281.32 \pm 6.82	2236.74 \pm 16.81	28.46 \pm 0.19
BiTGraph	6.13 \pm 0.01	10.76 \pm 0.02	12.41 \pm 0.12	270.14 \pm 3.77	2091.88 \pm 30.49	22.04 \pm 0.36
Method ($r = 0.6$)	MAE	Metr-LA RMSE	MAPE	MAE	Electricity RMSE	MAPE
BRITS	8.48 \pm 0.02	12.94 \pm 0.08	18.66 \pm 0.22	1029.38 \pm 1.84	10118.18 \pm 33.04	48.25 \pm 0.29
SPIN	6.61 \pm 0.02	11.35 \pm 0.17	13.31 \pm 0.12	—	—	—
GRIN	7.04 \pm 0.04	12.71 \pm 0.14	17.04 \pm 0.03	—	—	—
GCN-M	7.27 \pm 0.02	11.55 \pm 0.02	16.42 \pm 0.03	—	—	—
CRUs	11.02 \pm 0.02	13.38 \pm 0.24	20.40 \pm 0.04	664.07 \pm 9.88	8126.82 \pm 59.42	31.44 \pm 0.45
AGCRN ₀	14.87 \pm 0.04	14.30 \pm 0.09	29.92 \pm 0.06	1945.61 \pm 6.38	13891.03 \pm 17.38	75.20 \pm 0.29
Transformer ₀	7.46 \pm 0.01	12.03 \pm 0.05	17.09 \pm 0.07	346.43 \pm 5.59	2952.28 \pm 25.54	28.96 \pm 0.37
FEDformer ₀	7.50 \pm 0.04	12.32 \pm 0.03	17.31 \pm 0.07	535.72 \pm 7.67	5329.18 \pm 26.71	42.09 \pm 0.46
STWA ₀	6.82 \pm 0.02	11.72 \pm 0.10	13.66 \pm 0.04	325.47 \pm 6.62	2479.75 \pm 21.17	30.06 \pm 0.22
MTGNN ₀	6.95 \pm 0.03	12.09 \pm 0.02	13.87 \pm 0.09	329.18 \pm 4.61	2490.45 \pm 23.38	28.20 \pm 0.27
AGCRN _t	12.73 \pm 0.02	12.49 \pm 0.14	24.13 \pm 0.16	1374.64 \pm 7.11	12069.56 \pm 19.73	61.92 \pm 0.28
Transformer _t	7.22 \pm 0.08	13.61 \pm 0.17	16.75 \pm 0.06	327.17 \pm 8.68	2506.62 \pm 23.17	29.27 \pm 0.40
FEDformer _t	7.26 \pm 0.04	13.08 \pm 0.07	17.16 \pm 0.03	341.66 \pm 5.25	2682.73 \pm 24.97	29.87 \pm 0.09
STWA _t	6.55 \pm 0.02	11.28 \pm 0.07	13.57 \pm 0.03	312.25 \pm 5.36	2407.39 \pm 23.05	29.05 \pm 0.11
MTGNN _t	6.63 \pm 0.02	11.10 \pm 0.04	13.48 \pm 0.05	309.59 \pm 4.73	2399.51 \pm 20.09	25.37 \pm 0.13
BiTGraph	6.32 \pm 0.01	10.93 \pm 0.03	12.67 \pm 0.11	295.23 \pm 2.75	2239.06 \pm 26.39	27.38 \pm 0.49

Hình 5.3: Kết quả đầu ra của 2 bộ dữ liệu với độ thiếu dữ liệu là 0.1, 0.2, 0.4 và 0.6

Method ($r = 0.8$)	MAE	Metr-LA RMSE	MAPE	MAE	Electricity RMSE	MAPE
BRITS	8.56 ± 0.09	13.03 ± 0.18	18.92 ± 0.09	1027.28 ± 0.50	10150.54 ± 31.05	48.04 ± 0.02
SPIN	6.68 ± 0.31	11.42 ± 0.35	14.41 ± 1.20	—	—	—
GRIN	8.00 ± 0.02	12.68 ± 0.09	18.35 ± 0.05	—	—	—
GCN-M	7.75 ± 0.03	11.65 ± 0.04	17.94 ± 0.02	—	—	—
CRUs	11.35 ± 0.12	14.06 ± 0.70	22.08 ± 0.22	623.63 ± 13.07	7033.29 ± 17.85	33.29 ± 0.74
AGCRN ₀	14.86 ± 0.01	14.27 ± 0.02	29.92 ± 0.08	2351.41 ± 26.79	16824.28 ± 29.33	207.77 ± 0.56
Transformer ₀	8.06 ± 0.02	12.82 ± 0.05	18.37 ± 0.11	398.99 ± 6.62	3612.37 ± 24.19	30.07 ± 0.18
FEDformer ₀	7.83 ± 0.05	12.97 ± 0.14	17.93 ± 0.06	676.93 ± 5.62	7859.76 ± 31.13	64.79 ± 0.35
STWA ₀	7.57 ± 0.06	12.15 ± 0.07	17.31 ± 0.12	376.26 ± 5.36	3512.37 ± 22.09	31.15 ± 0.08
MTGNN ₀	7.45 ± 0.03	12.21 ± 0.08	17.22 ± 0.09	383.89 ± 6.72	3539.74 ± 15.22	30.29 ± 0.11
AGCRN _t	14.88 ± 0.01	14.20 ± 0.05	29.92 ± 0.10	1841.76 ± 6.87	17376.51 ± 44.79	70.38 ± 0.56
Transformer _t	7.32 ± 0.04	12.96 ± 0.08	16.87 ± 0.05	391.83 ± 4.17	3451.33 ± 5.62	32.26 ± 0.17
FEDformer _t	7.33 ± 0.06	13.17 ± 0.06	16.71 ± 0.04	380.06 ± 3.39	3335.18 ± 20.10	31.56 ± 0.13
STWA _t	6.90 ± 0.03	11.30 ± 0.05	13.69 ± 0.07	362.25 ± 3.21	3156.68 ± 24.41	29.22 ± 0.15
MTGNN _t	6.79 ± 0.04	11.05 ± 0.07	13.54 ± 0.10	355.68 ± 5.11	3023.30 ± 11.46	28.78 ± 0.31
BiTGraph	6.63 ± 0.01	<u>11.20 ± 0.00</u>	13.44 ± 0.02	347.35 ± 1.76	2839.79 ± 25.49	27.97 ± 0.27

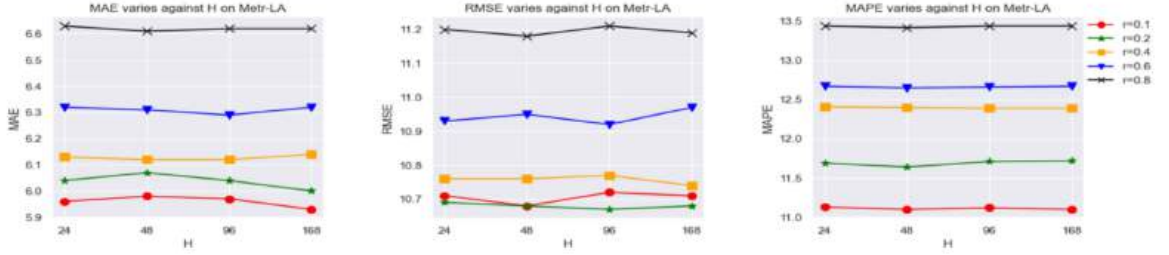
Hình 5.4: Kết quả đầu ra của 2 bộ dữ liệu với độ thiếu dữ liệu là 0.8

Kết luận

Trong nghiên cứu này, nhóm tác giả đã giới thiệu BiTGraph dành cho bài toán dự báo chuỗi thời gian với các giá trị bị thiếu. BiTGraph cùng lúc nắm bắt được động lực thời gian và cấu trúc không gian bằng cách xem xét một cách rõ ràng các giá trị bị thiếu. Họ đưa vào độ lệch (bias) trong hai mô-đun được thiết kế tỉ mỉ, đó là Multi-Scale Instance PartialTCN và Biased GCN, để tính đến các mẫu bị thiếu. Các kết quả thực nghiệm trên năm bộ dữ liệu tiêu chuẩn trong thế giới thực đã xác minh tính vượt trội của nó trong các tình huống giá trị bị thiếu khác nhau. Các nghiên cứu so sánh cũng cho thấy hiệu suất tuyệt vời của nó bắt nguồn từ hai thành phần Multi-Scale Instance PartialTCN và Biased GCN được thiết kế cẩn thận. Trong tương lai, nhóm nghiên cứu muốn khám phá kiến trúc Transformer như là kiến trúc chính của mô-đun thời gian để tăng cường hơn nữa hiệu suất dự báo dài hạn của nó cho dữ liệu chuỗi thời gian quan sát một phần.

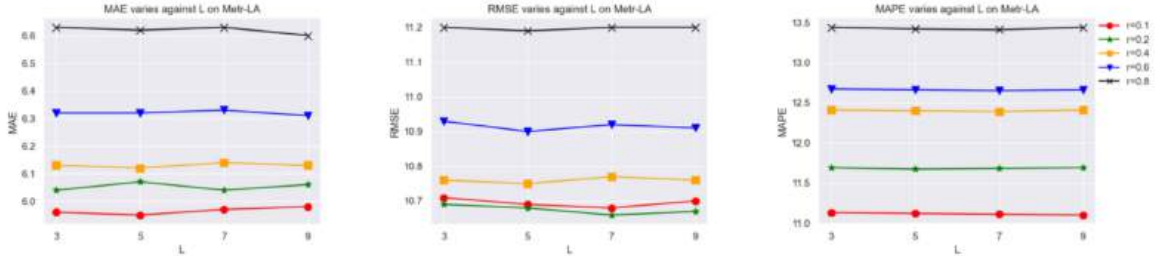
Thông tin thêm

7.1. Độ nhạy cảm của HYPERPARAMETER



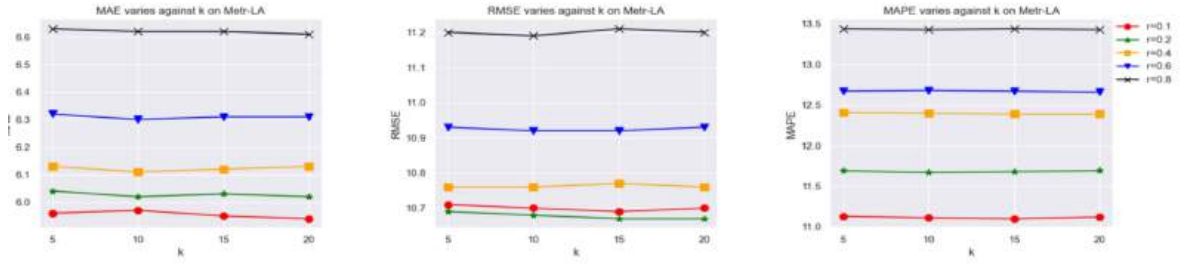
Hình 7.1: Kết quả với từng giá trị ví dụ của H .

- Hình 7.1 cho thấy sự thay đổi hiệu suất theo kích thước của sổ lịch sử H . Không có cải thiện hiệu suất rõ ràng nào được quan sát khi H tăng từ 24 lên 168.



Hình 7.2: Kết quả với từng giá trị ví dụ của L .

- Hình 7.2, BiTGraph đạt được hiệu suất tốt nhất khi $L = 5$. Tuy nhiên, hiệu suất không cải thiện thêm khi L tiếp tục tăng. Nhóm tác giả đưa ra giả thuyết rằng điều này là do vấn đề làm mịn quá mức (over-smooth) của GNN, nghĩa là, các biểu diễn nút trở nên không thể phân biệt được khi lớp tích chập đồ thị đạt đến 5.
- Sự thay đổi của các số liệu khi k thay đổi từ 5, 10, 15 đến 20. Như được quan sát từ hình 7.3, hiệu suất tăng lên bằng cách kết hợp thông tin không gian từ nhiều nút lân



Hình 7.3: Kết quả với từng giá trị ví dụ của k .

cận hơn. Tuy nhiên, một khi đủ các mối tương quan không gian được thu thập (tức là, $k=10$), việc tăng thêm số lượng hàng xóm sẽ không còn mang lại sự cải thiện hiệu suất. Điều này cho thấy rằng chỉ có một vài nút lân cận có tác động đáng kể đến một nút nhất định.

7.2. Vai trò của β

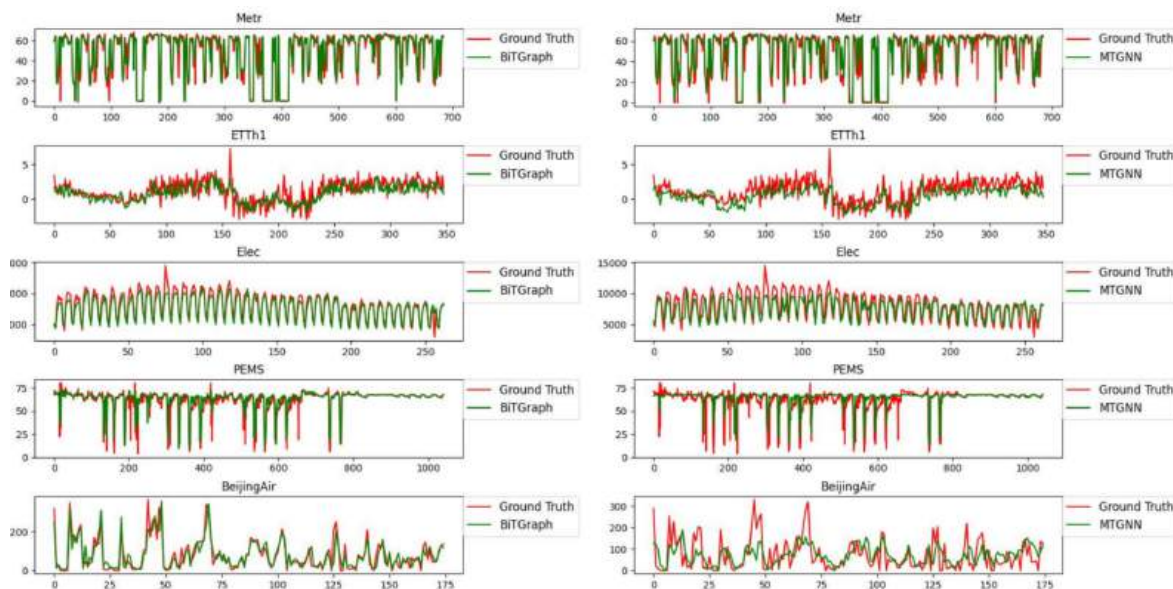
Dataset	0.2	0.4	0.6	0.8
Metr-LA	0.035	0.037	0.034	0.035
BeijingAir	0.475	0.463	0.471	0.467
ETTh1	0.397	0.391	0.419	0.419

Hình 7.4: Giá trị tương ứng của β trong từng trường hợp.

Hình 7.4 thể hiện sự thay đổi của β có thể học được theo tỷ lệ thiếu trên ba tập dữ liệu đại diện (giao thông, năng lượng mặt trời và chất lượng không khí). Nhóm nghiên cứu nhận thấy rằng β thể hiện các giá trị tương tự với các tỷ lệ thiếu khác nhau, cho thấy rằng β chủ yếu chịu trách nhiệm điều chỉnh độ mạnh của sự chính xác và độc lập với tỷ lệ thiếu. Ngoài ra, lý do có thể cho sự thay đổi giá trị của β trên các tập dữ liệu khác nhau có thể là do các đặc điểm riêng biệt của từng tập dữ liệu.

7.3. Trực quan kết quả giữa BiTGraph với MTGNN

Với tỉ lệ thiếu là 0.8, chúng ta dễ dàng nhận thấy tỉ lệ xuất hiện giá trị màu đỏ (tức là giá trị thực) xuất hiện ít hơn so với giá trị màu xanh (giá trị được dự đoán), càng nhiều giá trị màu đỏ bị đè bởi những đường màu xanh, có nghĩa là việc dự đoán dữ liệu ngày càng chính xác. Điều này thể hiện rằng kết quả dự đoán tại mô hình BiTGraph có độ chính xác cao và hiệu quả hơn so với mô hình MTGNN.



Hình 7.5: Kết quả BiGraph so với MTGNN.

7.4. Độ phức tạp của mô hình

Method	Memory Usage	#Parameters
BRITS	85.48M	173.72K
SPIN	10.85G	1.31K
GRIN	2.00G	12.76K
GCN-M	15.39G	396.50K
CRUs	41.97M	54.75K
AGCRN	10.17M	1.56K
MTGNN	250.55M	9.184K
Transformer	14.85G	15.46M
FEDformer	15.57G	16.08M
STWA	1.96G	256.50K
BiTGraph	194.17M	11.82K

Hình 7.6: Tài nguyên sử dụng của BiTGraph so với các phương pháp khác.

Nếu chỉ xét về riêng về độ tốn kém trong bộ nhớ khi sử dụng thì AGCRN sẽ là phương pháp có ưu thế, còn nếu xét về độ tốn kém trong việc sử dụng bộ nhớ thì SPIN là phương pháp tối ưu nhất. Tuy nhiên khi xét cả 2 vấn đề trên, AGCRN rất ấn tượng trong việc sử dụng và quản lý tài nguyên ở mức tối ưu nhất. BiTGraph sẽ là đại diện thứ 2 trong việc tối ưu tốt cả 2 vấn đề vừa rồi.

Nguồn tham khảo

References

- [1] Lei Bai et al. “Adaptive graph convolutional recurrent network for traffic forecasting”. In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2020.
- [2] Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. “An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling”. In: *arXiv preprint arXiv:1803.01271* (2018).
- [3] Wei Cao et al. “Brits: Bidirectional recurrent imputation for time series”. In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2018.
- [4] Andrea Cini, Ivan Marisca, and Cesare Alippi. “Filling the gaps: Multivariate time series imputation by graph neural networks”. In: *International Conference on Learning Representations (ICLR)*. 2022.
- [5] Razvan-Gabriel Cirstea et al. “Towards spatiotemporal aware traffic time series forecasting”. In: *International Conference on Data Engineering (ICDE)*. 2022.
- [6] Guilin Liu et al. “Image inpainting for irregular holes using partial convolutions”. In: *European Conference on Computer Vision*. 2018.
- [7] van Marisca, Andrea Cini, and Cesare Alippi. “Learning to reconstruct missing data from spatiotemporal graphs with sparse observations”. In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2022.
- [8] Mona Schirmer et al. “Modeling irregular time series with continuous recurrent units”. In: *International Conference on Machine Learning (ICML)*. 2022.
- [9] Haixu Wu et al. “TimesNet: Temporal 2D-Variation Modeling for General Time Series Analysis”. In: *International Conference on Learning Representations (ICLR)*. 2023.
- [10] Zonghan Wu et al. “Connecting the dots: Multivariate time series forecasting with graph neural networks”. In: *Proceedings of the 26th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. ACM. 2020, pp. 753–763.

- [11] George Zerveas et al. “A transformer-based framework for multivariate time series representation learning”. In: *ACM SIGKDD Conference on Knowledge Discovery Data Mining (SIGKDD)*. 2021.
- [12] Tian Zhou et al. “Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting”. In: *International Conference on Machine Learning (ICML)*. 2022.
- [13] Jingwei Zuo et al. “Graph convolutional networks for traffic forecasting with missing values”. In: *Data Mining and Knowledge Discovery* (2023).