

Quantificação de Incerteza em Redes Neurais em Grafos Aplicada à Previsão de Séries Temporais Multivariadas: Um Estudo de Caso em Tráfego Urbano

Ioná Santana

Orientador: Prof. Dr. Zanoni Dias

Coorientador: Prof. Dr. Hélio Pedrini

08 de Maio de 2024

Instituto de Computação
Universidade Estadual de Campinas

Introdução

Conceitos

Trabalhos Relacionados

Materiais e Métodos

Próximas Etapas

Introdução

Conceitos

Trabalhos Relacionados

Materiais e Métodos

Próximas Etapas

- O congestionamento de tráfego afeta a qualidade de vida e a segurança viária.
- Prever o fluxo de tráfego é útil para [1, 2]:
 - Otimização de rotas
 - Aumento da eficiência operacional
 - Melhoria da segurança viária
- Problemas na previsão [3, 4, 5, 6]:
 - Dados heterogêneos
 - Valores faltantes
 - Condições externas

Caracterização do Problema

- Sistemas de navegação e aplicativos de trânsito dependem dessas previsões para fornecer orientações precisas em tempo real
- As abordagens de previsão de tráfego negligencia a incerteza associada às previsões, limitando sua utilidade prática [7]
- A complexidade e interconexão dos sistemas urbanos desafiam os métodos tradicionais de previsão [8]

Desenvolver uma abordagem capaz de quantificar incertezas na previsão de séries temporais em redes neurais em grafos

Hipóteses do Trabalho

- Modelo capaz de generalizar previsões de fluxo de tráfego para diversas cidades urbanas
- Robustez testada em bases de dados temporais variadas, incluindo diferentes períodos
- Incorporação de técnicas de quantificação de incerteza para previsões mais confiáveis
- Análise das representações aprendidas revelando padrões e correlações não identificáveis por métodos tradicionais
- Desempenho computacional eficiente e escalável mesmo com grandes volumes de dados de tráfego urbano

Introdução

Conceitos

Trabalhos Relacionados

Materiais e Métodos

Próximas Etapas

Sequência de observações ou medições coletadas em intervalos sucessivos ao longo do tempo [9]

- **Série Temporal Univariada:** conjunto de observações coletadas ao longo do tempo para uma variável
- **Séries Temporais Multivariadas:** conjunto de observações coletadas ao longo do tempo para múltiplas variáveis simultaneamente

Redes Neurais em Grafo (GNNs)

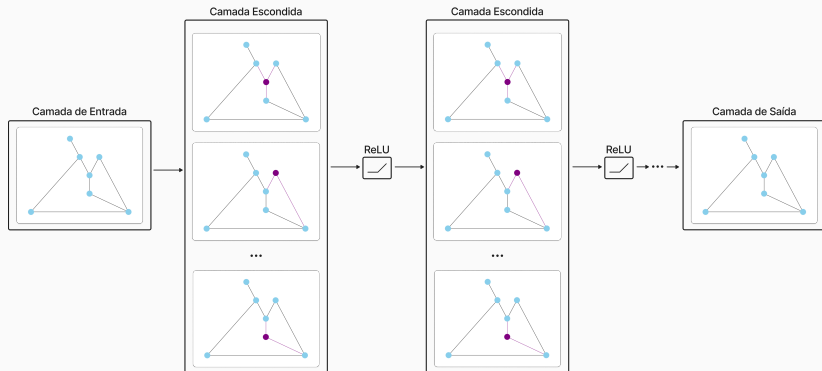


Figura 1: Funcionamento de uma GNN.

Representação Espaço-Temporal em GNNs (STGNNs)

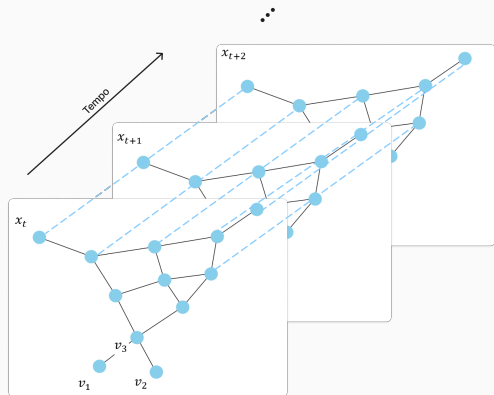


Figura 2: Representação Espaço-Temporal em GNNs.

Redes Neurais em Grafo (GNNs)

- No contexto do trabalho, os vértices do grafo são os sensores e a distância entre um sensor e outro são as arestas.

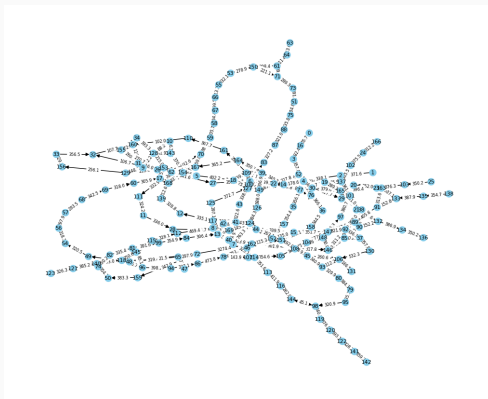
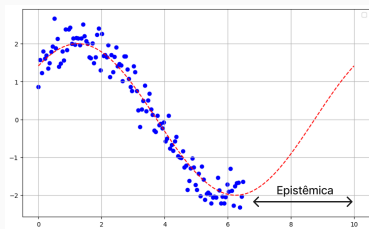


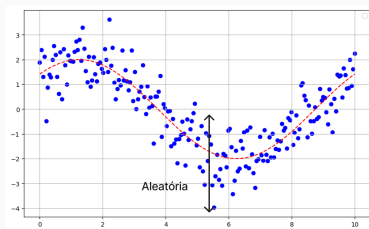
Figura 3: Sensores da base PeMSD8.

Quantificação de Incerteza (QI)

- Promover estimativas probabilísticas confiáveis.
- Lidar com incertezas inerentes aos dados.



(a) Incerteza Epistêmica.



(b) Incerteza Aleatória.

Figura 4: Quantificação de incertezas [10].

Introdução

Conceitos

Trabalhos Relacionados

Materiais e Métodos

Próximas Etapas

- GNNSTs
- Previsão de Tráfego
- Quantificação de Incerteza

Capacidade de incorporar informações de vizinhança e estrutura de rede em suas previsões

- HiSTGNN [11]
 - Captura de relações espaciais e temporais em variáveis meteorológicas e regiões
 - Desempenho superior a CNNs e RNNs
- TISER-GCN [12]
 - Redução média do MSE em 16,3% em dados sísmicos
 - Resultados comparáveis com metade da entrada
- Problemas:
 - Escalabilidade de GNNs para grafos grandes
 - Interpretabilidade dos resultados

- Slimani et al. [13]
 - Abordagem multivariada usando MLP, supera métodos univariados SARIMA e SMOreg
 - Menor erro relativo absoluto de 0,57
- Trinh et al. [14]
 - Modelos de DL multivariados superam modelos univariados e de aprendizado de máquina (ARIMA, Prophet, LSTM, TCN, Seq2Seq, NBeats)
- DGSTN [6]
 - Imputação de dados ausentes e captura de relacionamentos espaço-temporais
 - Resultados superiores a ARIMA, GCNs e métodos de atenção

Quantificação de Incerteza

- Processo Gaussiano de Substituição [15]
 - Aplicado a vários modelos bases
- RNNs e equações diferenciais estocásticas [16]
 - Imputação de séries temporais e a incerteza associada a essas imputações
- Regressão Linear Bayesiana [17]
 - Avaliação baseada em parâmetros de probabilidade a priori e a posterior
- DeepSTUQ [18]
 - Capacidade de estimar a incerteza epistêmica e aleatória
 - Bayesiana e ensembling

- Dependência de suposições específicas de dados, limitando a aplicabilidade do método
- Ineficiência computacional de algumas abordagens, exigindo múltiplos modelos adicionais para estimar a incerteza
- Incapacidade de alguns métodos em capturar com precisão a complexidade das correlações espaço-temporais nos dados
- Comunicação eficaz dos resultados

Introdução

Conceitos

Trabalhos Relacionados

Materiais e Métodos

Próximas Etapas

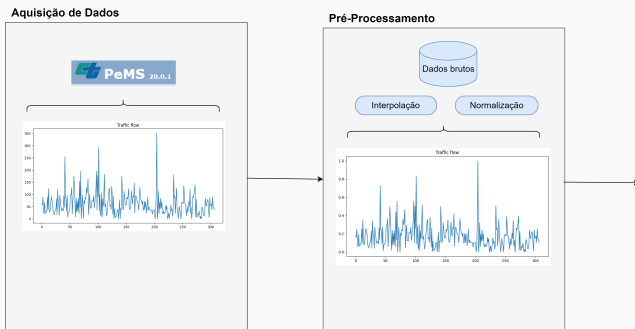


Figura 5: Visão geral do método proposto.

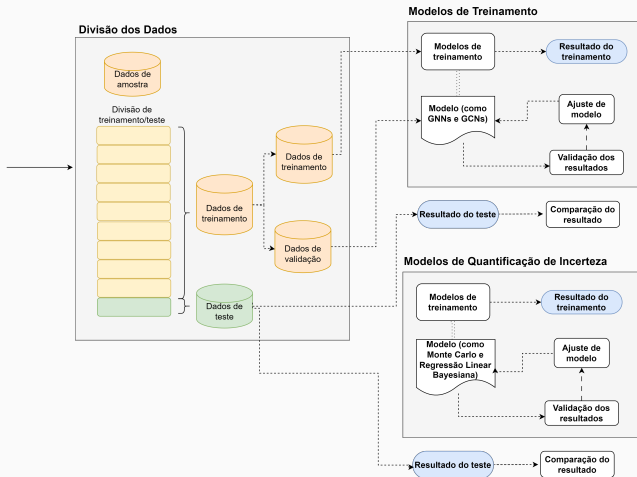


Figura 6: Visão geral do método proposto.

1. Aquisição de Dados

- Base PeMS

2. Pré-Processamento

- Imputação de dados faltantes
- Normalização dos dados
- Filtragem

3. Divisão dos Dados

- 90% para treinamento/validação e 10% para teste
- Método de validação cruzada em janelas deslizantes (*Time Series Split*)

4. Modelos para Previsão de Tráfego

5. Modelos para Quantificação de Incerteza

- Conjuntos de dados *Caltrans Performance Measurement System* (PeMS)
- Variáveis das bases: fluxo, ocupação, velocidade
- Base PeMSD4, dados de tráfego da cidade de San Bernardino:
 - Período: 01/07/2016 - 31/08/2016
- Base PeMSD8, dados de tráfego na região da Baía de São Francisco:
 - Período: 01/01/2018 - 28/02/2018

Tabela 1: Descrição espaço-temporal das bases de dados.

	Sensores	Tempo	Intervalo
PeMSD4	307	5 min	16992
PeMSD8	170	5 min	17856

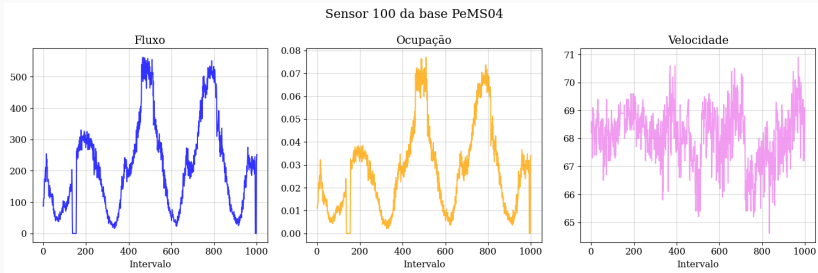
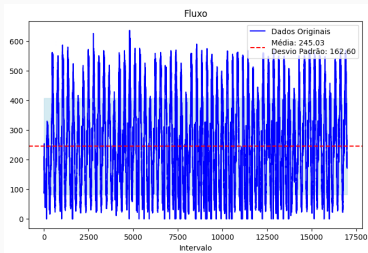
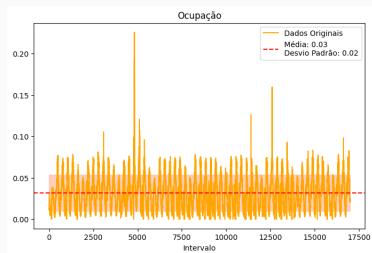


Figura 7: 1000 primeiros intervalos do fluxo, da ocupação e da velocidade do sensor 100 da base PEMSD4.



(a) Média e desvio padrão do fluxo do sensor 100.



(b) Média e desvio padrão da ocupação do sensor 100.

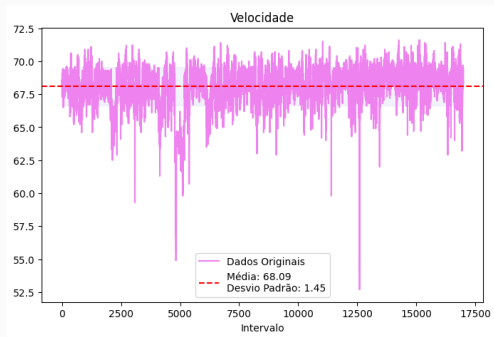


Figura 9: Média e desvio padrão da velocidade do sensor 100.

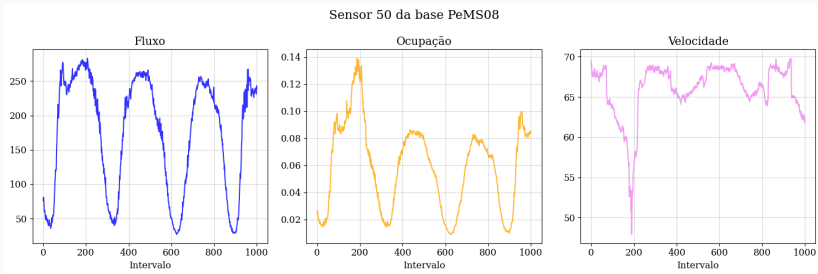
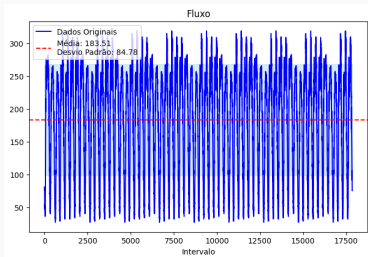
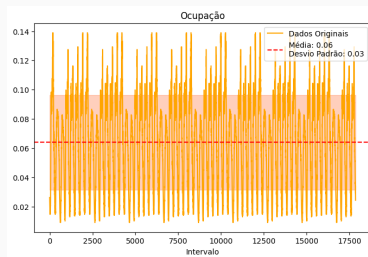


Figura 10: 1000 primeiros intervalos do fluxo, da ocupação e da velocidade do sensor 50 da base PEMSD8.



(a) Média e desvio padrão do fluxo do sensor 50.



(b) Média e desvio padrão da ocupação do sensor 50.

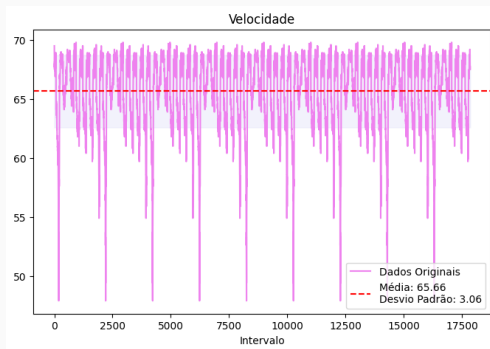


Figura 12: Média e desvio padrão da velocidade do sensor 50.

- Mean Absolute Error (MAE)
- Root Mean Squared Error (RMSE)
- Mean Absolute Percentage Error (MAPE)
- Empirical Correlation Coefficient (CORR)
- Incerteza Combinada (IC)

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (1)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (3)$$

$$\text{CORR} = \frac{\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2}} \quad (4)$$

$$\text{Aleatória} = E_{\theta \sim p(\theta)}[\sigma_{\theta}^2] \quad (5)$$

$$\text{Epistêmica} = V_{\theta \sim p(\theta)}[\mu_{\theta}] \quad (6)$$

$$\text{IC} = \sigma_{\text{Total}}^2 \approx \text{Aleatória} + \text{Epistêmica} \quad (7)$$

- Bibliotecas
 - NumPy
 - `scikit-learn`
 - TensorFlow
 - Keras
 - Matplotlib
- Ambiente virtual
 - Google Colaboratory

Plano de Trabalho e Cronograma de Execução

Atividades	1º ano						2º ano					
	1	2	3	4	5	6	1	2	3	4	5	6
Obtenção dos créditos obrigatórios do mestrado	•	•	•	•	•	•						
Pesquisa bibliográfica	•	•	•	•	•	•	•	•	•	•	•	•
Estudo e seleção de bases de dados		•	•	•	•							
Exame de Qualificação do Mestrado (EQM)				•								
Preparação da base de dados				•	•	•						
Implementação de um <i>baseline</i>					•	•	•					
Construção dos modelos de GNNs							•	•	•	•		
Construção dos modelos de QI								•	•	•	•	
Realização de testes								•	•	•	•	
Comparação dos resultados com outros trabalhos									•	•	•	
Publicação dos resultados										•	•	•
Escrita da dissertação							•	•	•	•	•	•
Defesa da dissertação												•

Introdução

Conceitos

Trabalhos Relacionados

Materiais e Métodos

Próximas Etapas

Próximas Etapas

- Preparação da Base de Dados:
 - Coleta e organização dos dados;
 - Tratamento de dados ausentes ou inconsistentes.
- Implementação de um Baseline:
 - Desenvolvimento de um modelo de referência;
 - Avaliação inicial do desempenho.
- Estudo de Abordagens com Transformers:
 - Revisão da literatura sobre *Transformers*;
 - Análise da aplicabilidade no contexto do trabalho.
- Potencial Uso de *Transformers*:
 - Exploração das capacidades para capturar dependências de longo alcance;
 - Investigação da viabilidade na previsão de séries temporais multivariadas.

- [1] Tomás Eloy Salais-Fierro and Jania Astrid Saucedo Martínez.
Demand forecasting for freight transport applying machine learning into the logistic distribution.
Mobile Networks and Applications, 27:2172–2181, 2022.
- [2] Yunduan Lin and Ruimin Li.
Real-time traffic accidents post-impact prediction: Based on crowdsourcing data.
Accident Analysis & Prevention, 145:105696, 2020.
- [3] Guohui Li, Haonan Deng, and Hong Yang.
Traffic flow prediction model based on improved variational mode decomposition and error correction.
Alexandria Engineering Journal, 76:361–389, 2023.

- [4] Wenbao Zeng, Ketong Wang, Jianghua Zhou, and Rongjun Cheng. **Traffic flow prediction based on hybrid deep learning models considering missing data and multiple factors.** *Sustainability*, 15(14):11092, 2023.
- [5] Meng Miao, Mingyu Kang, Xusheng Qian, Duxin Chen, Weijiang Wu, and Wenwu Yu. **Improving traffic time-series predictability by imputing continuous non-random missing data.** *IET Intelligent Transport Systems*, 17(10):1925–1934, 2023.
- [6] Ming Jiang and Zhiwei Liu. **Traffic flow prediction based on dynamic graph spatial-temporal neural network.** *Mathematics*, 11(11):2528, 2023.

- [7] M Jin, HY Koh, Q Wen, D Zambon, C Alippi, GI Webb, I King, and S Pan.

A survey on graph neural networks for time series: Forecasting, classification, imputation, and anomaly detection. arxiv 2023.

arXiv preprint arXiv:2307.03759.

- [8] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and S Yu Philip.

A comprehensive survey on graph neural networks.

IEEE Transactions on Neural Networks and Learning Systems, 32(1):4–24, 2020.

- [9] George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, and Greta M. Ljung.
Time Series Analysis: Forecasting and Control.
John Wiley & Sons, Inc., 5th edition, 2016.
- [10] Sibó Cheng, César Quilodrán-Casas, Said Ouala, Alban Farchi, Che Liu, Pierre Tando, Ronan Fablet, Didier Lucor, Bertrand Iooss, Julien Brajard, Dunhui Xiao, Tijana Janjic, Weiping Ding, Yike Guo, Alberto Carrassi, Marc Bocquet, and Rossella Arcucci.
Machine learning with data assimilation and uncertainty quantification for dynamical systems: A review.
IEEE/CAA Journal of Automatica Sinica, 10(6):1361–1387, 2023.

- [11] Minbo Ma, Peng Xie, Fei Teng, Bin Wang, Shenggong Ji, Junbo Zhang, and Tianrui Li.
HiSTGNN: Hierarchical spatio-temporal graph neural network for weather forecasting.
Information Sciences, 648:119580, 2023.
- [12] Stefan Bloemheuvel, Jurgen van den Hoogen, Dario Jozinović, Alberto Michelini, and Martin Atzmueller.
Graph neural networks for multivariate time series regression with application to seismic data.
International Journal of Data Science and Analytics, 16(3):317–332, 2023.

- [13] N Slimani, Ilham Slimani, N Sbiti, and M Amghar.
Machine learning and statistic predictive modeling for road traffic flow.
International Journal of Traffic and Transportation Management, 3(01):17–24, 2021.
- [14] Ngoc-Phap Trinh, Anh-Khoa N Tran, and Trong-Hop Do.
Traffic flow forecasting using multivariate time-series deep learning and distributed computing.
In *RIVF International Conference on Computing and Communication Technologies (RIVF)*, pages 665–670. IEEE, 2022.

- [15] Leonid Erlygin, Vladimir Zholobov, Valeriia Baklanova, Evgeny Sokolovskiy, and Alexey Zaytsev.
Uncertainty estimation for time series forecasting via Gaussian process regression surrogates.
arXiv preprint arXiv:2302.02834, 2023.
- [16] Shweta Dahale, Sai Munikoti, and Balasubramaniam Natarajan.
A General Framework for Uncertainty Quantification via Neural SDE-RNN.
arXiv preprint arXiv:2306.01189, 2023.
- [17] Muthyala Srinivas and P. Rama Chandra Prasad.
Modeling uncertainty quantification of ndvi of agricultural fields through bayesian linear regression in time series prediction.
Journal of Environmental Informatics Letters, 9(1):7–15, 2023.

- [18] W. Qian, D. Zhang, Y. Zhao, K. Zheng, and J. Q. Yu.
Uncertainty quantification for traffic forecasting: A unified approach.
In *IEEE 39th International Conference on Data Engineering (ICDE)*,
pages 992–1004, 2023.

Quantificação de Incerteza em Redes Neurais em Grafos Aplicada à Previsão de Séries Temporais Multivariadas: Um Estudo de Caso em Tráfego Urbano

Ioná Santana

Orientador: Prof. Dr. Zanoni Dias

Coorientador: Prof. Dr. Hélio Pedrini

08 de Maio de 2024

Instituto de Computação
Universidade Estadual de Campinas