



Universidade de Pernambuco (UPE)
Escola Politécnica de Pernambuco (POLI)
Instituto de Ciências Biológicas (ICB)

Coordenação de Pós-Graduação em Engenharia de Sistemas

Proposta de Dissertação de Mestrado

Área: Cibernética

Linha de Pesquisa: **Modelagem e simulação de sistemas inteligentes e embarcados**

Título Provisório: *Treinamento de modelos de imputação de dados faltantes em séries temporais multivariadas direcionado à detecção de anomalias.*

Orientador: Rodrigo de Paula Monteiro.

Co-orientador: Diego Marconi Pinheiro Ferreira Silva.

Descrição:

As cidades inteligentes produzem grandes volumes de dados [1] em domínios variados, como *Healthcare* [2] e Indústria 4.0 [3]. Essa abundância de dados se deve principalmente ao uso massivo de tecnologias de coleta de informação, ao acesso à internet banda larga e ao crescente poder computacional de dispositivos presentes no nosso cotidiano. A disponibilidade desses dados possibilita a extração de informações úteis na predição de eventos de interesse social, econômico, sanitário e ambiental [4].

Estes dados, normalmente coletados por sensores, podem ser representados como sequências de observações de diferentes variáveis realizadas ao longo do tempo, *i.e.*, séries temporais multivariadas. Tais séries estão presentes em diversos domínios, como: saúde, economia e computação. No entanto, incidentes inesperados podem provocar a ausência de valores nestas séries, *e.g.*, sensores quebrados, falha de transmissão, ou mesmo um simples esquecimento no preenchimento de uma planilha de acompanhamento [5]. Isto pode dificultar o entendimento dos fenômenos representados por meio deste tipo de informação, como o consumo energético de uma indústria ao longo dos anos, bem como a identificação de padrões em problemas de classificação, regressão ou previsão com dados de séries temporais. Isto torna a imputação, *i.e.*, a estimativa destes valores ausentes, um problema relevante na literatura, sendo um objeto de estudo presente numa quantidade significativa de trabalhos científicos [6].

Em trabalhos da literatura, observam-se diferentes formas de lidar com dados ausentes em séries temporais, múltiplas ou não. A forma mais simples e direta consiste na remoção de observações com valores ausentes em meio aos dados brutos. No entanto, métodos baseados em aprendizagem profunda, como as Redes Neurais Recorrentes (RNN, do inglês *Recurrent Neural Network*), as Memórias de Longo e Curto Prazo (LSTM, do inglês *Long Short-Term Memory*) e os Transformers vêm ganhando o protagonismo em problemas



Universidade de Pernambuco (UPE)
Escola Politécnica de Pernambuco (POLI)
Instituto de Ciências Biológicas (ICB)

relacionados à estimativa de valores faltantes em séries temporais, com uma ou mais variáveis.

Apesar do destaque de abordagens baseadas em aprendizagem profunda na estimativa de valores faltantes em séries temporais, ainda existem questões a serem desenvolvidas na área, principalmente relacionadas ao aprendizado de características em séries temporais multivariadas direcionado à detecção de anomalias. Isto é relevante porque a imputação normalmente é realizada para que os dados sejam melhorados para uma tarefa subsequente, e.g. o diagnóstico de falhas de uma máquina. Neste contexto, propõe-se uma pesquisa para identificar formas de se treinar modelos de imputação de dados em séries temporais multivariadas de modo a otimizar a detecção de falhas em dispositivos industriais. Este estudo será desenvolvido no contexto de máquinas rotativas industriais. Desta forma, o aprendizado de padrões presentes em séries temporais definidas por sensores que realizam o monitoramento de grandezas como a temperatura, nível de irradiação solar, dentre outras [7] são fundamentais.

Referências:

- [1] GORTON, Ian, et al. Data-intensive computing in the 21st century. *Computer*, 2008, 41.4: 30-32.
- [2] YOON, Jinsung; ZAME, William R.; VAN DER SCHAAR, Mihaela. Estimating missing data in temporal data streams using multi-directional recurrent neural networks. *IEEE Transactions on Biomedical Engineering*, 2018, 66.5: 1477-1490.
- [3] SHAHZAD, Khurram; O'NILS, Mattias. Condition monitoring in industry 4.0-design challenges and possibilities: A case study. In: 2018 Workshop on Metrology for Industry 4.0 and IoT. IEEE, 2018. p. 101-106.
- [4] ARCADINHO, Samuel; MATEUS, Paulo. Time series imputation. *arXiv preprint arXiv:1903.09732*, 2019.
- [5] CHE, Zhengping, et al. Recurrent neural networks for multivariate time series with missing values. *Scientific reports*, 2018, 8.1: 6085.
- [6] FANG, Chenguang; WANG, Chen. Time series data imputation: A survey on deep learning approaches. *arXiv preprint arXiv:2011.11347*, 2020.
- [7] SILVA, Miguel EP; VELOSO, Bruno; GAMA, João. Predictive Maintenance, Adversarial Autoencoders and Explainability. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Cham: Springer Nature Switzerland, 2023. p. 260-275.