CERGY PARIS
UNIVERSITÉ

Titre de la thèse/Thesis title : Explainable deep learning models for time series analytics (Classification, Forecasting and Anomaly Detection)
Laboratoire d'accueil / Host Laboratory : ETIS (MIDI team) (https://www.etis-lab.fr/midi/)
Spécialité du doctorat préparé/ Ph.D degree in : Computer Science (Informatique)
Mots-clefs / Keywords : Multivariate Time Series, Deep Learning, Machine Learning, Explainability

Descriptif détaillé de la thèse / Job description :

Introduction and Background.

Multivariate time series[i] (MTS) analysis such as Classification, Forecasting, and Anomaly Detection are omnipresent problems in many scientific domains. In this context, several state-of-the-art solutions to these problems adopt deep learning architectures such as CNN [ea20] (Convolutional Neural Network), LSTM [LGYC20, Pha21], GRU [SZN+19], and Transformer [VSP+17] (attention mechanism).
We note that these techniques typically exhibit shaky performance according to the analyzed data [FMBM21, GZS+21]. Hence, in order to adopt these algorithms, users first require to have a profound understanding of the data, which is finally validated by the correct interpretation of the analysis results.

In the last two or three years, we have witnessed a consistent effort to provide automatic *explainability* in deep learning models, which is the capability to highlight the feature sub-spaces that drive an algorithm toward a certain decision. Several works have tried to adapt explainability techniques natively used for images to MTS analysis. In this sense, we note that such adaption does not properly address the requirements of temporal data interpretation.

Specifically, several techniques such as Temporal Saliency Map [PHC] and MTEXCNN [AGBS] can provide the user with the spatio-temporal feature regions that characterize the class assignment of the adopted model. On the other hand, we note that these solutions cannot provide a global temporal explanation, namely, they can solely explain instances that are classified in a fixed time window. In this sense, many applications as predictive maintenance, supply-chain management, medical prognosis or urban crime prediction, and others require fully-fledged temporal explanations that reveal interesting temporal relationships among multiple univariate random variables.

Objective.

Time series data analysis naturally aims to discover complex events dependencies, namely temporal and inter-variable relationships that occur globally, e.g., anomaly root causes, variation of repeated patterns that fully describe the evolving structure of the data [ZINK19].

To that extent, **the Ph.D. candidate** will work towards the proposal of new models that can address the requirements of MTS analysis and are capable to explain the produced decisions. In a cognitive sense, a model must explain phenomena by appealing to their temporal relations rather than providing independent and unrelated explanations in separated time frames (as it is the case of current solutions).

Proposed Methodology.

We aim to design solutions for deep learning models that are able to leverage Complex Networks learning (e.g., Long-Run Variance Decomposition Networks [ea18]) and causality models such as Dynamic Bayesian Networks [SVC21, Mor21] , which effectively model complex events relations through graph structures. Specifically, we want to study both ante-hoc (integrated into the model) and posthoc (applied to the final results) explainability. Furthermore, we aim to extensively study and assess the applicability of our solutions in different heterogeneous use cases which call for effective results interpretability of MTS analysis (e.g., remote sensing, supply chain, climate data, and healthcare).

Références bibliographiques / Bibliography

- **[AGBS]** Roy Assaf, Ioana Giurgiu, Frank Bagehorn, and Anika Schumann. MTEX-CNN: multivariatetime series explanations for predictions with convolutional neural networks. In 2019 IEEE ICDM.
- **[ea18]** Eroglu et al. Multiplex recurrence networks. Physical Review E, 97, 01 2018.
- **[ea20]** Hassan Ismail Fawaz et al. Inceptiontime: Finding alexnet for time series classification. Data Min. Knowl. Discov., 34(6):1936–1962, 2020.
- **[FMBM21]** Cynthia Freeman, Jonathan Merriman, Ian Beaver, and Abdullah Mueen. Experimental comparison and survey of twelve time series anomaly detection algorithms. J. Artif. Intell. Res.,72:849–899, 2021.
- **[GZS+21**] Astha Garg, Wenyu Zhang, Jules Samaran, Ramasamy Savitha, and Chuan-Sheng Foo. An evaluation of anomaly detection and diagnosis in multivariate time series. IEEE Transactions on Neural Networks and Learning Systems, pages 1–10, 2021.
- **[LGYC20]** Yeqi Liu, Chuanyang Gong, Ling Yang, and Yingyi Chen. DSTP-RNN: A dual-stage two-phase attention-based recurrent neural network for long-term and multivariate time series prediction. Expert Syst. Appl., 143, 2020.
- **[Mor21]** Raha et al. Moraffah. Causal inference for time series analysis: problems, methods and evaluation. Knowledge and Information Systems, 2021.
- **[Pha21]** Tuan Pham. Time-frequency time-space long short-term memory networks for image classification of histopathological tissue. Scientific Reports, 11, 07 2021.
- **[PHC]** Qingyi Pan, Wenbo Hu, and Ning Chen. Two birds with one stone: Series saliency for accurate and interpretable multivariate time series forecasting. In IJCAI 2021.
- **[SVC21]** Jorge L. Serras, Susana Vinga, and Alexandra M. Carvalho. Outlier detection for multivariate time series using dynamic bayesian networks. Applied Sciences, 11, 2021.
- **[SZN+19]** Ya Su, Youjian Zhao, Chenhao Niu, Rong Liu,Wei Sun, and Dan Pei. Robust anomaly detection for multivariate time series through stochastic recurrent neural network. In ACM SIGKDD,2019.
- **[VSP+17]** Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. Advances in neural information processing systems, 30, 2017.
- **[ZINK19]** Yan Zhu, Makoto Imamura, Daniel Nikovski, and Eamonn J. Keogh. Introducing time series chains: a new primitive for time series data mining. Knowl. Inf. Syst., 60(2):1135–1161, 2019.

Profil demandé / Preferred Qualification

The candidate must fit the following requirements:

- Master's degree in computer science or data science.
- Advanced programming skills in Python (C++/Java is a plus).
- Strong mathematical background, including Linear Algebra and Statistics.
- Research experience in Machine learning, Deep Learning and Data Mining.
- Fluency in written and spoken English is essential.

Financement :
Début du contrat: Octobre, 1, 2022
Salaire mensuel brut : ~1975€ (Des missions doctorales sont possibles + ~400 euros brut)

Direction de la thèse:/ Thesis Supervisor:
Prof. Vassilis Chrsitophides (https://www.etis-lab.fr/2022/01/13/vassilis-christophides/)
Vassilis.Christophides@ensea.fr

Encadrement de la thèse : co-directeur(s) et co-encadrant(s)
Co-encadrant : Michele Linardi MCF (https://www.etis-lab.fr/2022/01/12/michele-linardi/)
michele.linardi@cyu.fr

Applicants should contact via email V. Christophides (Vassilis.Christophides@ensea.fr) and M. Linardi (michele.linardi@cyu.fr) with:

• A full curriculum vitae, including a summary of previous research experience.
• A transcript of higher education records
• A one-page research statement discussing how the candidate's background fits the proposed topic
• Two support letters of persons that have worked with them.

The deadline of the application is: June 3rd, 2022 (11h59 pm AoE).

---

[i] Multivariate time series are sets of one-dimension variables (a.k.a metrics), namely sequences of real values recorded along time at a fixed interval rate.