COMPUTER AND CONTROL ENGINEERING

TIERRA - Multivariate time series representation learning for vehicle telematics data analysis

Funded By	TIERRA S.P.A. [P.iva/CF:09888530012]
Supervisor	CAGLIERO LUCA - luca.cagliero@polito.it
Contact	CAGLIERO LUCA - luca.cagliero@polito.it VASSIO LUCA - luca.vassio@polito.it
Context of the research activity	This PhD proposal aims to study new techniques for embedding multivariate time series, apply them to solve established downstream tasks, and leverage these solutions in Data Science pipelines to analyze vehicles' telematics data such as CAN Bus signals. Embeddings will not only capture the series' temporal properties but also their multi-dimensional relations. These models will be used to classify, segment, and cluster signals and to detect anomalies and communities for industrial vehicle usage.
	Multivariate time series data have peculiar properties related to their sequential and multi-faceted nature. Although state-of-the-art embedding techniques tailored to time series data are effective in handling sequential data relations thanks to the use of auto-regressive or attention-based models, they often struggle to handle multiple dimensions at the same time. For example, CAN bus data acquired from vehicles cover a variety of different aspects (e.g., fuel level, coolant temperature, engine speed, ...) that are worth jointly analyzing to address predictive maintenance, anomaly detection, fleet detection and management, anomaly detection, and telematics service shaping. The PhD research will advance existing approaches to process and encode multivariate time series data, which encompass (but are not limited to) transformer models [1,2], contrastive and adversarial networks [3,4], matrix profile-based models [5,6], and Large Language Models [7]. The proposed representations will be then used to address various downstream tasks on time series data among which time series classification, forecasting, segmentation, and clustering and anomaly detection. For example, clustering and classifying CAN bus signals can be useful to automatically identify the working status of a vehicle according to both its performed activities and the environmental conditions [8]. Inter-series relations can be also analyzed to detect vehicle fleets and optimize resource allocation. Research objectives: Study of the state-of-the-art machine learning techniques for time series and compare their performance on the study case;

Objectives	Data collection and analysis of raw and structured data regarding vehicle telematics; Design, develop, test new approaches to time series representation; Benchmarking unimodal and multimodal time series models for time series classification, clustering, forecasting, and segmentation; Design new algorithms and methodologies to process time series data for supervised and unsupervised tasks. Industrial collaborations: The PhD activities will be supported by the ongoing research collaboration between Politecnico di Torino and Tierra Spa, a multinational telematics service provider that will provide in-domain data, expert supervision, and related case studies. In parallel, the research methods and algorithms can be also tested on benchmark data such as the UCR Time Series Classification Archive (https://www.cs.ucr.edu/~eamonn/time_series_data/) and mTAD (https://github.com/OpsPAI/MTAD). List of possible publication venues: - ECML PKDD, ACM CIKM, KDD, IEEE ICDE, IEEE ICDM, NEURIPS conferences - ACM TIST, ACM TKDD, IEEE TKDE, Elsevier Computers In Industry, Elsevier Information Sciences References: [1] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin: Attention is All you Need. NIPS 2017: 5998-6008 [2] Chao Yang, Xianzhi Wang, Lina Yao, Guodong Long, Guandong Xu: Dyformer: A dynamic transformer-based architecture for multivariate time series classification. Inf. Sci. 656: 119881 (2024) [3] Sana Tonekaboni, Danny Eytan, Anna Goldenberg: Unsupervised Representation Learning for Time Series with Temporal Neighborhood Coding. ICLR 2021 [4] hengyu Wang, Kui Wu, Tongqing Zhou, Zhiping Cai: Time2State: An Unsupervised Framework for Inferring the Latent States in Time Series Data. Proc. ACM Manag. Data 1(1): 17:1-17:18 (2023) [5] Eamonn J. Keogh: Time Series Data Mining: A Unifying View. Proc. VLDB Endow. 16(12): 3861-3863 (2023) [6] Yue Lu, Renjie Wu, Abdullah Mueen, Maria A. Zuluaga, Eamonn J. Keogh: DAMP: accurate time series anomaly detection on trillions of datapoints and ultra-fast arriving data streams. Data Min. Knowl. Discov. 37(2): 627-669 (2023) [7] Azul Garza, Max Mergenthaler Canseco: TimeGPT-1. CoRR abs/2310.03589 (2023) [8] Silvia Buccafusco, Luca Cagliero, Andrea Megaro, Francesco Vaccarino, Riccardo Loti, Lucia Salvatori: Learning industrial vehicles' duty patterns: A real case. Comput. Ind. 145: 103826 (2023)

Skills and competencies for the development of the activity	The PhD candidate is expected to - be proficient in Python and SQL programming - have a deep knowledge of statistics and probability fundamentals - have a solid background in data preprocessing and mining techniques - know the most established machine learning and deep learning techniques- have natural inclination for teamwork - be proficient in English speaking, reading, and writingWe seek motivated students who are willing to work at the intersection between academia and

	industry.