VRIJE UNIVERSITEIT BRUSSEL

The faculty of Engineering of the Vrije Universiteit Brussel invites you to attend the public defense leading to the degree of

DOCTOR OF ENGINEERING SCIENCES

of **Michiel Dhont**

The public defense will take place on **Thursday 3rd October 2024 at 4:00 pm** in room **I.2.02** (Building I, VUB Main Campus)

To join the digital defense, please click here

UNSUPERVISED ANALYTICS FOR MULTI-SOURCE TIME SERIES DATA

BOARD OF EXAMINERS

Prof. dr. ir. Heidi Ottevaere

Prof. dr. ir. Wendy Meulebroeck

Prof. dr. Lesley De Cruz

Prof. dr. ir. Bruno Cornelis

Prof. dr. ir. Dirk Deschrijver

Dr. ir. Wannes Meert

PROMOTORS

Dr. Elena Tsiporkova

Prof. dr. ir. Adrian Munteanu

VRIJE UNIVERSITEIT BRUSSEL

Abstract of the PhD research

There has been an explosion of impressive success stories recently with deep learning (DL) approaches in various fields such as natural language processing, computer vision, healthcare, and robotics. The advent of transformers has further amplified the capabilities of DL models to understand complex patterns, establishing them as a cornerstone of modern AI advancements across a broad spectrum of applications. Initially, transformers revolutionised large language models like GPT-4 and BERT, enabling them to process and generate human-like text with remarkable coherence and accuracy. Now, their impressive performance is also being demonstrated in other domains, extending their impact. Given sufficient high-quality labelled data and computational resources, DL models are able to achieve an accuracy that were previously unattainable.

Unfortunately, most of the real-world application contexts generate datasets which significantly diverge from the idealised benchmark datasets used to validate novel AI methodologies. Real-world data is typically characterised with presence of noise, missing values, complicated parameter names, different data types, lack of ground-truth, context-dependent features, etc. The latter makes it very challenging to immediately dive into any AI model application since it is often not clear which modelling paradigm best suit the problem at hand. This PhD research is built around the conception and validation of a heuristic data analytics methodology with the aim to benefit maximum from the different facets, while mitigating the imperfections, of real-world datasets.

Nowadays, most of the available datasets originating from industrial activities are composed of multitude of different parameters. The inherent multi-source nature of such datasets makes it impossible to directly integrate different data types without information loss. To address this challenge, a multi-view data integration approach has been devised as a part of this PhD, which identifies and considers different data views explicitly, allowing to fully harness the richness of heterogeneous datasets while retaining all relevant information.

The ongoing trend of increasingly more data being captured, goes parallel with an increasing complexity of extracting valuable insights from it. For instance, the remote monitoring of infrastructures (e.g., roads and power supplies) typically generates complex spatio-temporal data streams captured at high sampling rate across different locations. Combining and making sense of such data streams is not trivial. In this PhD research, a spatio-temporal profiling methodology is proposed, allowing to uncover insightful spatial patterns and dependencies while taking full advantage of the temporal dimension. Additionally, the exciting domain of visual analytics has been explored, resulting into the conception of several novel visualisation approaches, blending advanced visualisation with intelligent analysis to effectively reveal key iinsights.

By far, the hardest challenge associated with the analysis of real-world data is the lack of ground truth, which limits the choice of learning paradigms to only unsupervised ones. In this PhD research, a novel modelling framework is conceived, capable of extracting semantically interpretable states from unlabelled data. The latter facilitates a better understanding of system behaviour in terms of state transitions and allows to convert the unsupervised data modelling problem into a supervised one. Several different neural and neuro-symbolic forecasting workflows have been proposed for this purpose.