Explainability of time series prediction models for water resources management

The SIG Team of IRIT lab (UMR5505) of University Toulouse Capitole is offering an 12-month post-doctoral position in explainability as part of the ANR AIDA project which aims at proposing new actionable explainable models for the management of surface and sub-surface resources with a view to sustainable development.

The SIG team's research aims to design and develop methods, models, languages, algorithms and software tools that enable simple, effective access to relevant information, to improve its use, facilitate analysis and aid decision-making. The research work covers the entire data processing chain, from raw data to elaborate data accessible to users seeking information, wishing to visualize it and carry out decision-making, exploratory and predictive analyses.

Context

The AIDA project aims to address fundamental issues related to the quality and the applicability of explanations produced for Deep Learning (DL) models driven by the recent vision of **actionable explainable AI** (aXAI) [1]. AIDA considers so-called actionable concepts, measures, and metrics for explainable learning and reasoning. In particular, it focuses on **more expressive forms of explanations** that can answer not only *why* questions (why do we obtain a specific prediction, given the features of input observations?) but also action-guiding explanations such as *how-to* (what are the necessary actions to change the prediction of a specific input observation?) and *what-if* (what are the necessary and minimal sets of actions on input observations required to obtain an alternative prediction?). Answers to these questions are crucial to act on the models and the data used in various prediction tasks of real applications [2].

To ground our research program on aXAI, we focus on a water resources management use-case that is (i) of prime importance in the context of climate pledge at a national level and (ii) a core interest for the French Geological Survey BRGM. The maintenance and deployment of DL models for groundwater levels forecasting [3] fusing multivariate time series (MTS) at multiple regions of the French territory currently face several challenges (site equipment specificity, geological and anthropogenic context). To overcome those challenges, BRGM experts strive for explanations on which segment (e.g. sensor, site, model architecture, preprocessing steps) of the modeling procedure impacts the final result. This in turn will help them to identify any misuse of data by the model, or conversely will highlight new relations among observations that could potentially improve in-house BRGM models or ease their deployment on new sites.

Current XAI approaches suffer from two main limitations [4]. The first [5] is related to the complexity of the explanation process itself, which involves the particular characteristics of "the

training data, the precise shape of the decision surface, and the selection of one explanatory algorithm over another". As a result, there is a risk of accepting plausible explanations that only reflect spurious correlations between internal layers of DL with input features. [5] emphasizes the need for explanations methods "that cast doubt on certain features of AI systems". These results call for (i) a more thorough consideration of inner relationships in the data and how models use this information, (ii) methods to assist users in selecting explanation methods based on objective metrics. The second limitation is related to the lack "of XAI approaches tackling real-world machine learning issues" that would "help to clarify what is currently feasible and what is not feasible when employing XAI techniques" [4].

Objectives and tasks

During the postdoc we envision several research questions attached to the aforementioned objectives of quality and applicability of XAI approaches:

1. **Literature survey of quality metrics for causal explanations**. This work will be conducted in conjunction with another post-doc of the university of Tours who will be in charge of implementing a benchmark and a library of existing quality metrics from [6] for feature influence, rule-based and counterfactual explanations

2. Development of new **quality measures for causal explanations**. Several research directions are envisioned
   a. build new measures as crossover between existing metrics
   b. build new interestingness scores based on expert domain knowledge from BRGM
   c. ideally, organize user tests with BRGM experts to validate the proposed quality measures

3. **Exploration of the relationships between explanations** (possibly of different kinds). This last point will be optionally addressed only if time allows.
   a. Development of adequation measures to determine which XAI method (influence-based, rule-based or causal) is most appropriate based on the dataset and ML models to explain
   b. Complementarity of explanations: some predictions can be supported by several explanations either of different kinds or supporting each other.

Collaboration

The recruited candidate will be supervised by Julien Aligon and Moncef Garouani (IRIT, Université Toulouse Capitole) in partnership with Nicolas Labroche (Université of Tours), BRGM (the French Geological Survey) and the PhD student and postdoc already at work in the project to develop new ML models.
Other external collaborations will be conducted with XAI experts from other labs such as ETIS and LIP6 (Paris).

Skills / Qualifications

PhD in Computer Science, specializing in artificial intelligence (explainability, possibly deep learning), experience in processing temporal data (multivariate time series and multivariate event sequences). Experience working with libraries offering implementations of XAI and deep learning models. Experience in setting user protocol would be appreciated.

General information

Work location (in-office): the hired candidate will work in Toulouse, Université Toulouse Capitole, Manufacture des Tabacs, 21 Allée de Brienne 3100

Duration : 12 months, starting from March 2025
Gross salary : approx. 2500€/month

To apply, please email your CV, cover letter, list of publications, and references (including their email addresses, affiliations, and a brief description of your professional relationship with each) exclusively via email.

The deadline to apply is 10st January 2025, 12H00. Applications will be accepted until the position is filled.

Contact *Julien Aligon*, *Moncef Garouani, Nicolas Labroche and Alexandre Chanson* to obtain further information: julien.aligon@irit.fr, moncef.garouani@irit.fr, nicolas.labroche@univ-tours.fr, alexandre.chanson@univ-tours.fr

[1] xxAI - Beyond Explainable AI, Andreas Holzinger, Randy Goebel, Ruth Fong, Taesup Moon, Klaus-Robert Müller, Wojciech Samek - International Workshop, Held in Conjunction with ICML 2020, July 18, 2020, Vienna, Austria, Revised and Extended Papers. Lecture Notes in Computer Science 13200, Springer 2022, ISBN 978-3-031-04082-5.
[2] Causal Explanations and XAI, Sander Beckersn, 2022,1st Conference on Causal Learning and Reasoning, CLeaR 2022},Proceedings of Machine Learning Research.
[3] Un modèle hydrologique spatialisé pour la simulation de très grands bassins : le modèle EROS, forme de grappes de modèles globaux élémentaires, Thierry, D.. and Moutzopoulos, C.,1995, Régionalisation en hydrologie : application au développement.
[4] Why Is the Current XAI Not Meeting the Expectations?, Alessio Malizia and Fabio Paterno, 2023, Commun. ACM.
[5] Post-Hoc Explanations Fail to Achieve their Purpose in Adversarial Contexts, Bordt, Sebastian and Finck, Michele and Raidl, Eric and others, 2022, 2022 ACM Conference on Fairness, Accountability, and Transparency.
[6] From Anecdotal Evidence to Quantitative Evaluation Methods: A Systematic Review on Evaluating Explainable AI, Nauta, Meike and Trienes, Jan and Pathak, Shreyasi and others, 2023, ACM Computing Surveys.