TAILOR

Foundation of Trustworthy AI:
Integrating Learning, Optimisation and Reasoning

POSTER SESSION

2nd TAILOR conference, Prague 13-14 September 2022

Poster overview

Last name	First name	Organisation	Session
Baruchelli	Paola	FBK	2
Bitetto	Alessandro	University of Pavia	1
Cavalcante Siebert	Luciano	TU Delft	2
Giannini	Francesco	CINI	1
Harley	Nicholas	VUB	2
Hofmann	Till	RWTH Aachen University	1
Homola	Martin	Slovak.AI	2
Isaksson	Alf	ABB Corporate Research	2
Karystinaios	Emmanouil	Johannes Kepler University (JKU)	1
Kocev	Dragi	Jozef Stefan Institute	1
Koubarakis	Manolis	National and Kapodistrian University of Athens	1
Krawiec	Krzysztof	Poznan University of Technology (PUT)	2
Kučera	Petr	Charles University	1
Meyer-Vitali	André	DFKI	2
Ondrčková	Simona	Charles University	1
Orlandini	Andrea	National Research Council of Italy (CNR-ISTC)	2
Panov	Pance	Jozef Stefan Institute	1
Pecher	Branislav	Slovak.AI	1
Potoniec	Jedrzej	Poznan University of Technology	1
Pratesi	Francesca	ISTI - CNR	2
Riguzzi	Fabrizio	CINI - University of Ferrara	1
Sansone	Emanuele	KU Leuven	1
Sansone	Emanuele	KU Leuven	1
Schulz	Jasmin	Luxembourg Institute of Health	2
Shindo	Hikaru	TU Darmstadt	1
Skrzypczynski	Piotr	Poznan University of Technology	1
Stefanowski	Jerzy	Poznan University of Technology	1
Švancara	Jiří	Charles University	1
Szeląg	Marcin	Poznan University of Technology	1
Tamajka	Martin	slovak.AI	2
Teso	Stefano	University of Trento	1
Teso	Stefano	University of Trento	1
Treguer	Sebastien	INRIA	2
Yorke-Smith	Neil	TU Delft	1

Contacts and information:
Email: trine.vikinge@liu.se Website: www.tailor-network.eu

Session 1

Alessandro Bitetto, University of Pavia
Oracle-LSTM: a neural network approach to mixed frequency time series prediction

In the context of macro-economic indicators there are two main concerns regarding the frequency of the variables. The first is related to MIxed DAta Sampling (MIDAS), i.e. some indicators are reported annually, some quarterly, others monthly. The second deals with the need of forecasting predictions between reporting dates, e.g. before the end of the year, and it is known as "nowcasting". Existing methods rely on the alignment of high-frequency input data to low-frequency target variable by the means of lagged variables and their temporal-decaying weighting. We develop a two-steps algorithm that makes use of two Recurrent Neural Networks. The first, called Oracle, is a Deep Autoregressive network and predicts the target variable at high-frequency given past information. The second, called Predictor, is Long-Short Term Memory (LSTM) network and learns the relationship between Oracle's predictions and high-frequency input data. The prediction error is a weighted average of two terms: one compares the observed low-frequency target with predictions of both Oracle and Predictor, the other compares the Predictor's high-frequency predictions with the Oracle's ones. Our model is tested on both simulated data, where we know the generated high-frequency data, and real macro-economic data. Our results show better performances compared to classical approaches. Moreover, we apply gradient-based interpretability methods to estimate the input features' importance in the predictions.

Francesco Giannini, CINI
Relational Reasoning Networks

Authors: Giuseppe Marra, Michelangelo Diligenti, Francesco Giannini

Neuro-symbolic methods integrate neural architectures, knowledge representation and reasoning. However, they have been struggling at both dealing with the intrinsic uncertainty of the observations and scaling to real-world applications.
This poster presents Relational Reasoning Networks (R2N), an end-to-end model performing logic reasoning in the latent space of a deep learner architecture, where the embedding representations of constants, ground atoms and ground formulas are learned in an integrated fashion. In facts, we define a computational structure accounting for both relations between entities and logic formulas among ground atoms. R2Ns can be applied to purely symbolic tasks or as a neuro-symbolic platform to integrate learning and reasoning in heterogeneous problems with both symbolic and feature-based represented entities. The proposed model overtakes some limitations of previous neuro-symbolic methods and achieves state-of-the-art results in different experimental settings.

Till Hofmann, RWTH Aachen University
Controlling Golog Programs against MTL Constraints

Author: Till Hofmann

While Golog is an expressive programming language to control the high-level behavior of a robot, it is often tedious to use on a real robotic system. On an actual robot, the user needs to consider low-level details, such as enabling and disabling hardware components, e.g., a camera to detect objects for grasping. In other words, high-level actions usually pose implicit temporal constraints on the low-level platform, which are typically independent of the program to be executed. In this paper, we propose to make these constraints explicit by modeling them with Metric Temporal Logic, which allows to enforce the execution of certain low-level platform operations in addition to the main program. Based on results from timed automata controller synthesis, we describe a method to synthesize a controller that executes both the high-level program and the low-level platform operations concurrently in order to satisfy the MTL specification. This allows the user to focus on the high-level behavior without the need to consider low-level operations.

Emmanouil Karystinaios, Johannes Kepler University (JKU)
The "Whither Music?" Project

Authors: Gerhard Widmer, and Team

WHITHER MUSIC? is a project that aims to establish model-based computer simulation (via methods of AI, (deep) Machine Learning and probabilistic modelling) as a viable methodology for asking questions about musical processes, developments, possibilities and alternatives - for music research, for didactic purposes, for creative music exploration scenarios.

Dragi Kocev, Jozef Stefan Institute
Explaining the performance of multilabel classification methods with data set properties

Author: Dragi Kocev

Meta learning generalizes the empirical experience with different learning tasks and holds promise for providing important empirical insight into the behavior of machine learning algorithms. Here, we present a comprehensive meta-learning study of data sets and methods for multilabel classification (MLC). MLC is a practically relevant machine learning task where each example is labeled with multiple labels simultaneously. Here, we analyze 40 MLC data sets by using 50 meta features describing different properties of the data. The main findings of this study are as follows. First, the most prominent meta features that describe the space of MLC data sets are the ones assessing different aspects of the label space. Second, the meta models show that the most important meta features describe the label space, and, the meta features describing the relationships among the labels tend to occur a bit more often than the meta features describing the distributions between and within the individual labels. Third, the

optimization of the hyperparameters can improve the predictive performance, however, quite often the extent of the improvements does not always justify the resource utilization.

Manolis Koubarakis, National and Kapodistrian University of Athens
Efficient Learning of Multiple NLP Tasks via Collective Weight Factorization on BERT

The Transformer architecture continues to show remarkable performance gains in many Natural Language Processing tasks. However, obtaining such state-of-the-art performance in different tasks requires fine-tuning the same model separately for each task. Clearly, such an approach is demanding in terms of both memory requirements and computing power.
In this paper, aiming to improve training efficiency across multiple tasks, we propose to collectively factorize the weighs of the multi-head attention module of a pre-trained Transformer. We test our proposed method on finetuning multiple natural language understanding tasks by employing BERT-Large as an instantiation of the Transformer and the GLUE as the evaluation benchmark. Experimental results show that our method requires training and storing only 1% of the initial model parameters for each task and matches or improves the original fine-tuned model's performance for each task while effectively decreasing the parameter requirements by two orders of magnitude. Furthermore, compared to well-known adapter-based alternatives on the GLUE benchmark, our method consistently reaches the same levels of performance while requiring approximately four times fewer total and trainable parameters per task.
This paper has been published in NAACL2022. Full version at
https://cgi.di.uoa.gr/~koubarak/publications/2022/NAACL22_PAPER__all_.pdf
Joint work with Christos Papadopoulos, Yannis Panagakis and Mihalis Nikolaou.

Petr Kučera, Charles University
Propagation Complete Encodings as a Target Compilation Language

Author: Petr Kučera

We describe properties of propagation complete encodings as a target language in knowledge compilation. We also recollect the results on compiling a knowledge representation into a propagation complete encoding. The source compilation languages we consider include decomposable negation normal form (DNNF), backdoor decomposable monotone circuits (BDMC), and CNF formulas.

Simona Ondrčková, Charles University
Verification, recognition and correction of hierarchical plans.

Authors: Simona Ondrčková, Roman Barták

We will introduce our work in the areas of plan verification, plan recognition and plan and model correction in hierarchical planning. Hierarchical planning is a form of planning where tasks decompose into smaller easier tasks. Hierarchical plan verification focuses on checking whether an action sequence is a valid plan (is executable and a goal task can be decomposed into it). Plan

recognition focuses on predicting a plan based on a set of observed actions. Plan correction is about changing an invalid plan by adding or removing actions to make it valid. Finally model correction is about adjusting the model to make a given plan valid.

Pance Panov, Jozef Stefan Institute
FAIRification of MLC data

Authors: Ana Kostovska, Jasmin Bogatinovski, Andrej Treven, Sašo Džeroski, Dragi Kocev, and Panče Panov

The MLC task has increasingly been receiving interest from the machine learning (ML) community, as evidenced by the growing number of papers and methods that appear in the literature. Hence, ensuring proper, correct, robust, and trustworthy benchmarking is of utmost importance for the further development of the field. We believe that this can be achieved by adhering to the recently emerged data management standards, such as the FAIR (Findable, Accessible, Interoperable, and Reusable) and TRUST (Transparency, Responsibility, User focus, Sustainability, and Technology) principles. To FAIRify the MLC datasets, we introduce an ontology-based online catalogue of MLC datasets that follow these principles. The catalogue extensively describes many MLC datasets with comprehensible meta-features, MLC-specific semantic descriptions, and different data provenance information. The MLC data catalogue is available at: http://semantichub.ijs.si/MLCdatasets. In addition, we provide an ontology-based system for easy access and querying of performance/benchmark data obtained from a comprehensive MLC benchmark study. The system is available at: http://semantichub.ijs.si/MLCbenchmark.

Branislav Pecher, Slovak.AI
Stability of Learning With Limited Labelled Data

Authors: Branislav Pecher, Ivan Srba, Maria Bielikova

Learning with limited labelled data, such as meta-learning or transfer learning, can be used to achieve high performance when the available labels are lacking. However, the limited availability of labels makes the training process more unstable and prone to the effects of uncontrolled randomness. Even small changes in data, such as changing order of samples, may lead to massive changes in performance. When instability of the models is not explicitly taken into consideration, it may result in biased comparisons between approaches and lack of trust in behaviour of the models in practice. We want to remedy this by in-depth exploration of randomness factors that influence the stability of learning with limited labelled data. We design

a methodology for investigating these effects and their further mitigation to reduce the instability. Applying this methodology to meta-learning approaches on a sentiment detection task, we identify key randomness factors influencing the stability of selected meta-learning approaches.

Jedrzej Potoniec, Poznan University of Technology
Towards Learning Concept Embeddings with a Transferable Deep Neural Reasoner

Authors: Dariusz Max Adamski, Jedrzej Potoniec

We present a novel approach for learning embeddings concepts in knowledge bases expressed in the description logic ALC. The embeddings reflect the semantics of the concepts in such a way that it is possible to compute an embedding of a complex concept from the embeddings of its parts by using appropriate neural constructors. Embeddings for different knowledge bases are vectors in a shared vector space, shaped in such a way that approximate subsumption checking for arbitrarily complex concepts can be done by the same neural network, called a reasoner head, for all the knowledge bases. We report the results of experimental evaluation showing that the difference in reasoning performance between training a separate reasoner head for each ontology and using a shared reasoner head, is negligible.

Fabrizio Riguzzi, CINI - University of Ferrara
Statistical Statements in Probabilistic Logic Programming

Authors: Damiano Azzolini, Elena Bellodi, Fabrizio Riguzzi

Probabilistic Logic Programs under the distribution semantics (PLPDS) do not allow statistical probabilistic statements of the form ``90\% of birds fly'', which were defined ``Type 1'' statements by Halpern.
In this paper, we add this kind of statements to PLPDS and introduce the PASTA (``Probabilistic Answer set programming for STAtistical probabilities'') language. We translate programs in our new formalism into probabilistic answer set programs under the credal semantics. This approach differs from previous proposals, such as the one based on ``probabilistic conditionals'' as, instead of choosing a single model by making the maximum entropy assumption, we take into consideration all models and we assign probability intervals to queries. In this way we refrain from making assumptions and we obtain a more neutral framework.
We also propose an inference algorithm and compare it with an existing solver for probabilistic answer set programs on a number of programs of increasing size, showing that our solution is faster and can deal with larger instances.

Emanuele Sansone, KU Leuven
LSB: Local Self-Balancing MCMC in Discrete Spaces

Author: Emanuele Sansone

We present the Local Self-Balancing sampler (LSB), a local Markov Chain Monte Carlo (MCMC) method for sampling in purely discrete domains, which is able to autonomously adapt to the target distribution and to reduce the number of target evaluations required to converge. LSB is based on (i) a parametrization of locally balanced proposals, (ii) a newly proposed objective function based on mutual information and (iii) a self-balancing learning procedure, which minimises the proposed objective to update the proposal parameters. Experiments on energy-based models and Markov networks show that LSB converges using a smaller number of queries to the oracle distribution compared to recent local MCMC samplers.

Emanuele Sansone, KU Leuven
Recent Advances on Neuro-Symbolic Learning at KU Leuven

Authors: Robin Manhaeve, Emanuele Sansone

In this poster, we present recent work on neuro-symbolic learning performed at KU Leuven. Specifically, we discuss three research lines, namely the integration between neural networks and stochastic logic programs, an approximate inference strategy for probabilistic logic programming and a generative neuro-symbolic framework based on variational autoencoders.

Hikaru Shindo, TU Darmstadt
alphaILP: Thinking Visual Scenes as Differentiable Logic Programs

Authors: Hikaru Shindo, Viktor Pfanschilling, Devendra Singh Dhami, Kristian Kersting

Deep neural learning has shown remarkable performance at learning representations for visual object categorization. However, deep neural networks such as CNNs do not explicitly encode objects and relations among them. This limits their success on tasks that require a deep logical understanding of visual scenes, such as Kandinsky patterns and Bongard problems. To overcome these limitations, we introduce alphaILP, a novel differentiable inductive logic programming framework that learns to represent scenes as logic programs—intuitively, logical atoms correspond to objects, attributes, and relations, and clauses encode highlevel scene information. alphaILP has an end-to-end reasoning architecture from visual inputs. Using it, alphaILP performs differentiable inductive logic programming on complex visual scenes, i.e., the logical rules are learned by gradient descent. Our extensive experiments on Kandinsky patterns and CLEVR-Hans benchmarks demonstrate the accuracy and efficiency of alphaILP on learning complex visual-logical concepts.

Piotr Skrzypczynski, Poznan University of Technology
Geometry-Aware Keypoint Network: Accurate Prediction of Point Features with Spatial Uncertainty

Authors: Piotr Skrzypczynski, Tomasz Nowak

We consider a challenging scenario of localising a camera with respect to a charging station for electric buses. In this application, we face a number of problems, including a substantial scale change as the bus approaches the station, and the need to detect keypoints on a weakly textured object in a wide range of lighting and weather conditions. Therefore, we use a deep convolutional neural network to detect the features, while retaining a conventional procedure for pose estimation with 2D-to-3D associations. We leverage here the backbone of HRNet, a state-of-the-art network used for detection of feature points in human pose recognition, and we further improve the solution adding constraints that stem from the known scene geometry. We incorporate the reprojection-based geometric priors in a novel loss function for HRNet training and use the object geometry to construct sanity checks in postprocessing. Moreover, we demonstrate that our Geometry-Aware Keypoint Network yields feasible estimates of the geometric (spatial) uncertainty of point features.

Jerzy Stefanowski, Poznan University of Technology
Increasing the interpretability of rules induced from imbalanced data.

Author: Jerzy Stefanowski

Although classification rules are usually considered as a directly interpretable representation of knowledge induced from data, sometimes the algorithms may induce too many rules for human inspection. In particular, it occurs for algorithms designed for improving classification of imbalanced data, such as e.g. BRACID. Therefore the analysts could be supported in the process of identifying the limited subset of the most interesting rules. For this aim we will present a multicriteria approach, where various rule evaluation measures are exploited. In our experiment we consider specialized Bayesian confirmation measures combined with the measure of the rule support. Experimental results show that this approach can reduce the number of rules induced by BRACID algorithm and improve values of rule interestingness measures at the same time, without considerable losses of classification accuracy, especially for the minority class.

Jiří Švancara, Charles University
On Solving Multi-agent Pathfinding

Author: Jiří Švancara

Multi-agent pathfinding is the problem of navigating a set of agents in a shared environment such that they do not collide with each other. This problem has numerous practical applications such as warehousing, intersection management, railway navigation, game development, and many more. In this poster, we present our experience with solving multi-agent pathfinding and some of its uses.

Marcin Szeląg, Poznan University of Technology
Explainable rule models in analysis of classification data with monotonicity constraints

Authors: Marcin Szeląg, Roman Słowiński

We present a framework and software for learning decision rule models from data describing an ordinal classification of objects. It employs Dominance-based Rough Set Approach that handles partially inconsistent and incomplete data where attributes are of heterogeneous nature – cardinal, ordinal, and nominal. Considering the dominance relation among objects, the data are pre-processed by calculating lower approximations of sets of objects belonging to ordered decision classes. Then, these approximations are used to induce monotonic decision rules. Each rule presents a scenario of a causal relationship between evaluation of objects on considered attributes and a classification decision. They do not only constitute a glass-box data model but can also be employed to classify new objects using different classification strategies. Then, the matching rules constitute readable arguments for particular class assignments.

Stefano Teso, University of Trento
Concept-level Debugging of Part-Prototype Networks

Author: Stefano Teso

Part-prototype Networks (ProtoPNets) are concept-based classifiers designed to achieve the same performance as black-box models without compromising transparency. ProtoPNets compute predictions based on similarity to class-specific part-prototypes learned to recognize parts of training examples, making it easy to faithfully determine what examples are responsible for any target prediction and why. However, like other models, they are prone to picking up confounds and shortcuts from the data, thus suffering from compromised prediction accuracy and limited generalization. We propose ProtoPDebug, an effective concept-level debugger for ProtoPNets in which a human supervisor, guided by the model's explanations, supplies feedback in the form of what part-prototypes must be forgotten or kept, and the model is fine-tuned to align with this supervision. An extensive empirical evaluation on synthetic and real-world data shows that ProtoPDebug outperforms state-of-the-art debuggers for a fraction of the annotation cost.

Stefano Teso, University of Trento
GlanceNets: Interpretabile, Leak-proof Concept-based Models

Author: Stefano Teso

There is growing interest in concept-based models (CBMs) that combine high-performance and interpretability by acquiring and reasoning with a vocabulary of high-level concepts. A key requirement is that the concepts be interpretable. Existing CBMs tackle this desideratum using a variety of heuristics based on unclear notions of interpretability, and fail to acquire concepts with the intended semantics. We address this by providing a clear definition of interpretability in terms of alignment between the model's representation and an underlying data generation process, and introduce GlanceNets, a new CBM that exploits techniques from disentangled representation learning and open-set recognition to achieve alignment, thus improving the

interpretability of the learned concepts. We show that GlanceNets, paired with concept-level supervision, achieve better alignment than state-of-the-art approaches while preventing spurious information from unintendedly leaking into the learned concepts

Neil Yorke-Smith, TU Delft
Agent-Based Simulation of Short-Term Peer-to-Peer Rentals: Evidence from the Amsterdam Housing Market

Author: Neil Yorke-Smith

We study the effect of a range of possible municipal policy measures on the peer-to-peer short-term rental market. The case study is the city of Amsterdam. A spatial agent-based simulation indicates that more lower income citizens remain in the city centre when regulation of the market is stronger, and that banning the touristic market restrains the overall increase in house prices, compared to the business-as-usual scenario. However, the feasibility of enforcement of regulation, and its libertarian consequences, must be considered. The full article is published in Environment and Planning B https://doi.org/10.1177%2F23998083211000747

Session 2

Paola Baruchelli, FBK
Trustworthy AI

Author: Paola Baruchelli

The poster will introduce FBK approach to Trustworthy AI in line with the work done on WP8.

Luciano Cavalcante Siebert, TU Delft
Meaningful human control: actionable properties for AI system development

Authors: Luciano Cavalcante Siebert (presenter), Maria Luce Lupetti, Evgeni Aizenberg, Niek Beckers, Arkady Zgonnikov, Herman Veluwenkamp, David Abbink, Elisa Giaccardi, Geert-Jan Houben, Catholijn M. Jonker, Jeroen van den Hoven, Deborah Forster & Reginald L. Lagendijk

How can humans remain in control of artificial intelligence (AI)-based systems designed to perform tasks autonomously? Such systems are increasingly ubiquitous, creating benefits - but also undesirable situations where moral responsibility for their actions cannot be properly attributed to any particular person or group. The concept of meaningful human control has been proposed to address responsibility gaps and mitigate them by establishing conditions that

enable a proper attribution of responsibility for humans; however, clear requirements for researchers, designers, and engineers are yet inexistent, making the development of AI-based systems that remain under meaningful human control challenging. In this paper, we address the gap between philosophical theory and engineering practice by identifying, through an iterative process of abductive thinking, four actionable properties for AI-based systems under meaningful human control, which we discuss making use of two applications scenarios: automated vehicles and AI-based hiring. First, a system in which humans and AI algorithms interact should have an explicitly defined domain of morally loaded situations within which the system ought to operate. Second, humans and AI agents within the system should have appropriate and mutually compatible representations. Third, responsibility attributed to a human should be commensurate with that human's ability and authority to control the system. Fourth, there should be explicit links between the actions of the AI agents and actions of humans who are aware of their moral responsibility. We argue that these four properties will support practically minded professionals to take concrete steps toward designing and engineering for AI systems that facilitate meaningful human control

Nicholas Harley, VUB
CHAKRA: Common Hierarchical Abstract Knowledge Representation for Anything

Author: Nicholas Harley

Research in many fields is increasingly collaborative and data driven. Sharing and reuse of digital research knowledge requires generalised representations which afford uniform access, manipulation and reasoning across application contexts and domains. CHAKRA is a knowledge representation framework which allows for the integration and logical description of distributed, heterogeneous research data. This poster illustrates the core components of the framework.

Martin Homola, Slovak.AI
Explainable Malware Detection

Authors: Martin Homola, Ján Kľuka, Alexander Šimko, Daniel Trizna, Peter Švec, Štefan Balogh, Umberto Straccia (and others)

The coordinated action "Explainable Malware Detection" under the frame of WP3.1 aims to investigate XAI methods that can be used to explain outcomes of malware detection methods. To this end either XAI methods can be adopted and applied on top of MD tools, or MD methods can be enhanced and become more explainable. Under our CA we have so far focused on concept learning which has been applied on top of EMBER - which is a pre-classified malware dataset. One of our main results is a handcrafted ontology that can be paired with EMBER and is needed for this task. We are working with a number of concept learning toolkits (including DL learner, DL FOIL/FOCL, Fuzzy DL learner) which we hope to evaluate and compare on this use case. In the following period we would like to extend our investigation also on KB Embedding and Ontology extraction from neural classifiers. We hope to compare the applicability of these methods to the state-of-the art concept learning.

Alf Isaksson, ABB Corporate Research
Industrial AI at ABB Corporate Research in Västerås, Sweden & Ladenburg, Germany

Authors: Alf Isaksson, Martin Hoffmann, Shiva Sander Tavallaey and Benjamin Klöpper

The poster describes the importance of AI for ABB, a global leader in Automation and Electrification. In particular the need for explainable AI is elaborated. The poster also expresses a call for visits by other TAILOR researchers to our ABB Corporate Research Centers in Sweden and Germany.

Krzysztof Krawiec, Poznan University of Technology (PUT)
Counterexample-Driven Genetic Programming for Symbolic Regression with Formal Constraints

Authors: Iwo Błądek and Krzysztof Krawiec

In symbolic regression with formal constraints, the conventional formulation of regression problem is extended with desired properties of the target model, like symmetry, monotonicity, or convexity. We present a genetic programming algorithm that solves such problems using a Satisfiability Modulo Theories solver to formally verify the candidate solutions. The essence of the method consists in collecting the counterexamples resulting from model verification and using them to improve search guidance. The method is exact: upon successful termination, the produced model is guaranteed to meet the specified constraints. We compare the effectiveness of the proposed method with standard constraint-agnostic machine learning regression algorithms on a range of benchmarks, and demonstrate that it outperforms them on several performance indicators.

André Meyer-Vitali, DFKI
Trustworthy Hybrid Team Decision-Support

Authors: André Meyer-Vitali, Wico Mulder

The aim to empower human users of artificially intelligent systems becomes paramount when considering coordination in hybrid teams of humans and autonomous agents. Hereby, we consider not only one-to-one interactions, but also many-to-many situations (multiple humans and multiple agents), where we strive to make use of their complementary capabilities. Therefore, mutual awareness of each others' strengths and weaknesses is crucial for beneficial coordination. In order to address these goals, and in accordance with a hybrid theory of mind, we propose the use of trustworthy interaction patterns and epistemic orchestration with intentions and causal models.

Andrea Orlandini, National Research Council of Italy (CNR-ISTC)

Deploying Dependable AI System: Barriers and Solutions

Authors : Andrea Orlandini, Alessandro Umbrico, Amedeo Cesta

This poster describes some research activities performed by the Planning and Scheduling Technology (PST) Laboratory at Institute of Cognitive Science and Technology of National Research Council of Italy (CNR-ISTC) to deal with dependability issues of AI applications. Specifically, we discuss barriers using human-robot collaboration applications in manufacturing as a reference scenario.

Francesca Pratesi, ISTI - CNR

Comparison of methods for interpretation of convolutional neural networks for the classification of multiparametric MRI images on unbalanced datasets

Authors: Plamen Angelov, Andrea Berti, Franco Alberto Cardillo, Sara Colantonio, Franca Debole, Elisa Fromont, Luis Galarraga, Fosca Giannotti, Oana Goga, Riccardo Guidotti, Patrick Loiseau, Carlo Metta, Marah Mohammad, Anna Monreale, Eva Pachetti, Cecilia Panigutti, Francesca Pratesi, Georges Quenot, Hossein Rahmani, Salvatore Rinzivillo, Marie-Christine Rousset, Umberto Straccia, Romain Xu-Darme

Within the TAILOR task 3.1 - Explainable AI Systems, we would like to apply several methods to explain the decisions of neural network in a same specific case study, i.e., classification of multiparametric Magnetic Resonance Imaging scans of prostates of patience having cancers, a case study characterized by a large unbalancing of the data.
Starting from a existing neural network, we aim at: (i) improving the results obtained by the given classifier; (ii) testing different algorithms to provide explanations, both from the state of the art and designed by the TAILOR partners; (iii) comparing and generalizing the adopted AI solutions, possibly creating guidelines for a standardised process.

Jasmin Schulz, Luxembourg Institute of Health

Clinnova - unlocking the potential of data science and AI in healthcare

Authors: Jasmin Schulz

AI-driven solutions for precision medicine have still not yet lived up to expectations – and their potential. Effective AI solutions require a data-enabling environment, which is based both on infrastructure investment and coordination between clinical players. Clinnova is an international project with the aim to support the digitalisation of healthcare and precision medicine by creating a data-enabling environment for accessing, sharing and analysing interoperable, high-quality health data. Patient data will be analysed in a federated network of local health data hubs, rendering the creation of a framework of trusted and explainable essential for success.

Martin Tamajka, slovak.AI

Optimizing Post-hoc Explainability Algorithm for Finding Faithful and Understandable Explanations for a Combination of Model, Task and Data

Author: Martin Tamajka

We address the problem of explaining a specific prediction of a model. First, we propose a method for finding a configuration of a post-hoc explainability algorithm, Layer-wise relevance propagation (LRP), that provides understandable and faithful explanations for a particular model, task, and data. The method is based on modified Particle Swarm Optimization. We propose a novel fitness function, which in combination with a fidelity measure Area over perturbation curve (AOPC) is used to focus the relevance assigned by the LRP algorithm to expected regions. This makes the explanations both faithful and understandable for humans. Second, we generalize the proposed method to an arbitrary set of post-hoc explainability algorithms. We evaluated the proposed method on two image classification tasks (magnetic resonance imaging datasets BraTS and ADNI) and one text sentiment classification task (dataset SST). Our method outperformed benchmark LRP configurations in all three tasks, which suggests its extensibility to different architectures, datasets and tasks.

Sebastien Treguer, INRIA

Learning To Run a Power Network Challenge: Energies of the Future and Carbon Neutrality – 2022

Authors : Gaetan Serré, Sébastien Treguer, Marc Schoenaeur

Current rapid changes in climate increase the urgency to change energy production and consumption management, to reduce carbon and other green-house gas production. In this context, the French electricity network management company RTE (Réseau de Transport d'Électricité) has recently published the results of an extensive study outlining various scenarios for tomorrow's French power management. We propose a challenge that will test the viability of such a scenario. The goal is to control electricity transportation in power networks, while pursuing multiple objectives: balancing production and consumption, minimizing energetic losses, and keeping people and equipment safe and particularly avoiding catastrophic failures. While the importance of the application provides a goal in itself, this challenge also aims to push the state-of-the-art in a branch of Artificial Intelligence (AI) called Reinforcement Learning (RL), which offers new possibilities to tackle control problems. In particular, various aspects of the combination of Deep Learning and RL called Deep Reinforcement Learning remain to be harnessed in this application domain. This challenge belongs to a series started in 2019 under the name "Learning to run a power network" (L2RPN). In this new edition, we introduce new more realistic scenarios proposed by RTE to reach carbon neutrality by 2050, retiring fossil fuel electricity production, increasing proportions of renewable and nuclear energy and introducing batteries. Furthermore, we provide a baseline using state-of-the-art reinforcement learning algorithm to stimulate the future participants.