



ReaLearn: Reasoning and Learning for Trustworthy AI

研究を始めるのに必要な知識・能力

We welcome students who want to build AI systems that humans can trust! **Preferred:** Knowledge of elementary set theory and/or basic statistics and/or computer programming skills (e.g. Python, Java) are helpful to work on this field.

この研究で身につく能力

Students will be trained to broaden their knowledge of AI covering both (1) Knowledge Representation and Reasoning (KRR) and (2) Machine Learning (ML), and then apply the knowledge to develop Trustable AI systems. They will learn to work with different forms of input based on their interest. For example, tabular data, Knowledge Graph, Ontology, and unstructured data (e.g. text, images, and videos). Eventually, they will develop Trustworthy AI algorithms that enable to provide fair, explainable and/or robust prediction. Improving Transparency, Fairness, and Accountability in addition to AI performance is the ultimate goal of this lab. This skill is essential for making AI systems in today's era.

【就職先企業・職種】 Software Industry, AI Industry

研究内容

Our research laboratory aims to address research questions lied in between pure (logic) reasoning and (pure) machine learning for Trustworthy AI, especially Explainable AI (XAI).

Our study covers **two mainstreams of AI**, i.e., (1) Knowledge Representation and Reasoning (**KRR**) and (2) Machine Learning (**ML**).

The research themes are divided into 3 sub-themes under the umbrella of Trustworthy AI and XAI.

Theme 1: KRR and Explanation Formalism

Since the 90s, explanation in KRR is studied to support users for various purposes, e.g., for explaining logical reasoning or debugging knowledgebases. However, most explanations are still difficult to understand by humans and are merely available for specific logical reasoning algorithms. We are attempting this issue on the following KRR formalisms (but not limited to):

- Description Logic, Ontology, Knowledge Graph
- Argumentation and Explanation from Proof
- Virtual Knowledge Graph

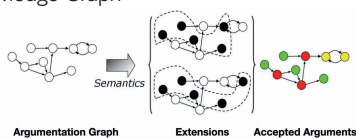


Figure 1 Argumentation and its Semantics

Theme 2: Fair and Explainable ML

We study rigorous computer science techniques necessary for fair and explainable ML in addition to the performance of ML algorithms. This theme is socially important and has strong industrial interest due to the increasing regulations around the world. We are attempting this issue on the following areas:

- Natural Language Processing
- Computer Vision

Theme 3: Integration of KRR and ML for Trustworthy AI

Combining aspects of KRR and ML has received a great deal of attention in recent years, either in the form of interaction between both paradigms or in the form of hybrid systems combining both components into one architecture, such as,

- Argument Mining (i.e. Argumentation meets Machine Learning)
- Knowledge Graph Embedding (i.e. Knowledge Graph meets Machine Learning)
- Ontology Embedding (i.e. Ontology meets Machine Learning)
- Neural-Symbolic Learning (i.e. Symbolic Reasoning meets Machine Learning)

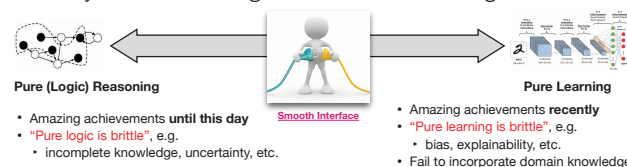


Figure 2 Integration of KRR and ML. Why should we integrate them?

主な研究業績

1. Teeradaj Racharak, Interpretable Decision Tree Ensemble Learning with Abstract Argumentation for Binary Classification, ICONIP 2022
2. Teeradaj Racharak, Doing Analogical Reasoning in Dynamic Assumption-based Argumentation Frameworks, ICTAI 2022
3. Wei Kun Kong, Teeradaj Racharak, Yiming Cao, Cheng Peng, Minh Le Nguyen, KGWE: A Knowledge-guided Word Embedding Fine-tuning Model, ICTAI 2021

使用装置

CPU/GPU Cluster Machines

研究室の指導方針

To enhance student's abilities, I conduct:

1. Intensive discussion on one-to-one meetings,
2. Research progress and group discussion on lab meeting,
3. Guidance on improving scientific paper writing and presentation skills,
4. Reading groups to enhance research thinking.

In addition, I will support students to broaden their network with leading experts in respective fields.

【研究室HP】 URL : <https://www.jaist.ac.jp/~racharak/lab>