Jinqiang Yu

PhD Candidate at Monash University

Wellington Rd, Clayton VIC 3800, Australia

✉ jinqiang.yuuu@gmail.com | in linkedin.com/in/jinqiang-yu-404bb0187 | 🎓 Jinqiang Yu

Summary

I am passionate about machine learning and AI areas, with experience in the fields of machine/deep learning and explainable AI, which is a division of responsible AI. As an enthusiast of cutting-edge AI techniques, I particularly has a strong desire to explore them, such as generative AI and large language models. With the enthusiasm and experience in machine learning and AI areas, I am expecting to utilise my strong technical skills to contribute to real-world projects and acquire valuable industry experience.

Education

Monash University *Melbourne, Australia*

PhD in Data Science & AI *Feb 2021 - Jul 2024*

- **Thesis Topic**: Explainable AI with the Use of Formal Reasoning
- **Supervisors**: Prof. Peter J. Stuckey, Dr. Alexey Ignatiev
- **Thesis Description**: In this PhD project, we aim to develop approaches to generating interpretable machine learning models, e.g. decision trees/ sets/ lists, as well as devise approaches to accurately and concisely explaining predictions made by machine/deep learning models. Our research tackles explainability challenges in various domains, including NLP, image classification, and other general classification tasks.

Monash University *Melbourne, Australia*

Master of Information Technology *Mar 2019 - Dec 2020*

- Graduated with H1
- **Core units**: Master Minor Thesis, Applied Data Analysis, Data Processing for Big Data, Statistical Data Modelling, Algorithm and Data Structure.
- **Minor Thesis Topic**: Computing optimal interpretable machine learning models.
- **Thesis Description**: The thesis focuses on the interpretable models, e.g. decision trees/ sets/ lists, aiming at developing advanced approaches to computing machine learning models that are both accurate and interpretable.

Publications

From Formal Boosted Tree Explanations to Interpretable Rule Sets
Jinqiang Yu, Alexey Ignatiev, Peter J. Stuckey
29th International Conference on Principles and Practice of Constraint Programming (CP), 2023

On Formal Feature Attribution and Its Approximation
Jinqiang Yu, Alexey Ignatiev, Peter J. Stuckey
arXiv preprint arXiv:2307.03380, 2023

Eliminating the Impossible, Whatever Remains Must Be True: On Extracting and Applying Background Knowledge in the Context of Formal Explanations
Jinqiang Yu, Alexey Ignatiev, Peter J. Stuckey, Nina Narodytska, Joao Marques-Silva
37th AAAI Conference on Artificial Intelligence (AAAI), 2023

Learning Optimal Decision Sets and Lists with SAT
Jinqiang Yu, Alexey Ignatiev, Peter J. Stuckey, Pierre Le Bodic
Journal of Artificial Intelligence Research (JAIR) 72 (2021) pp. 1251–1279. 2021

Computing Optimal Decision Sets with SAT
Jinqiang Yu, Alexey Ignatiev, Peter J. Stuckey, Pierre Le Bodic
26th International Conference on Principles and Practice of Constraint Programming (CP), 2020

Research Projects

Explaining and Correcting LLM Outputs *Aug 2023 - current*

- Recently, Satisfiability-Aided language modeling (SATLM) was introduced to generate a declarative task specification rather than an imperative program and leverage an off-the-shelf automated theorem prover to derive the final answer. Although SATLM improves the reasoning capabilities of LLMs, there is still lack of explainability of why the outputs are correct/incorrect. Inspired by this observation, we aim at developing the approach to explaining which outputs of LLMs trigger incorrect answers and correcting the outputs to get correct answers from the solver.
- **Technical Skills:** Python, OpenAI, PySAT.

Explainability in NLP and Image Classification *Dec 2022 - present*

- As ML models for NLP/image classification problems are black-box models, users cannot understand the prediction made by the models and thus it is hard to trust the predictions. Although existing model-agnostic approaches are able to provide explainability for predictions, these approaches are known to suffer from fundamental explanation issues. Inspired by the limitation, in this project, we target developing the approach to providing trustable explanations for NLP/image predictions in machine/deep learning models.
- **Technical Skills:** Python, Tensorflow, XGBoost, PyTorch, PySAT.

Applying Trustable Explanations in Real-world Scenarios *Mar 2022 - present*

- Due to the lack of explainability of ML and AI, humans cannot understand the reason behind the predictions made by ML models. For example, in the context of Just-In-Time defect prediction, practitioners cannot know why a commit is predicted as defect-introducing because of the lack of explainability of defect models. Motivated by the limitation, this project aims at developing approaches to applying explainable AI in practical scenarios such that users can trust the predictions made by ML models and also find an explainable way to change the decision.
- **Technical Skills:** Python, XGBoost, scikit-learn, PySAT.

Computing Succinct and Accurate Explanations *Feb 2021 - present*

- In recent years the growing practical AI and ML applications have given the rise to Explainable AI (XAI). One of the major approaches to XAI is to compute explanations to ML predictions on demand, including post-hoc (*abductive*) explanations answering a "*why?*" question and (*contrastive*) explanations targeting a "*why not?*" question. In this project, we focus on developing the approach to computing both abductive and contrastive formal explanations making use of background knowledge, which can positively affect the quality of both kinds of explanations.
- **Technical Skills:** Python, PyTorch, XGBoost, scikit-learn, PySAT.

Learning Optimal Interpretable Machine Learning Models *Feb 2020 - present*

- In order to make explanations easy for humans to understand the interpretable models, e.g. decision trees, lists and sets, they should be as concise as possible. In addition, such models should provide accurate predictions such that humans can make proper decisions based on the predictions. Therefore, this project focuses on devising approaches to computing interpretable ML models that are both small in size and accurate, making use of modern formal reasoning.
- **Technical Skills:** Python, XGBoost, scikit-learn, PySAT.

Experience

Optima *Melbourne, Australia*

Student Researcher *Apr 2021 - Present*

- Under the supervision of Prof. Peter J. Stuckey from Optima, I am engaged in a PhD project at Monash University. My research focuses on explainable AI, including generating interpretable ML models and computing accurate and concise explanations, aiming at developing methods and techniques to help users understand and explain ML model inferences and predictions.

Monash University *Melbourne, Australia*

Teaching Associate *Jun 2021 - Nov 2021*

- Tutoring and grading students in tutorials, assignments, and final exams.
- **Unit**: FIT5220 - Solving discrete optimisation problems.
- **Topics**: Constraint Programming, Mixed Integer Programming, Boolean Satisfiability (SAT) Solving, Large Neighbourhood Search.

Skills

Proficient in Python, Java, R, SQL.

Familiar with C/C++, Spark, MongoDB, MATLAB, machine/deep learning models (neural networks, transformers, LLMs, GANs, and gradient boosting trees, etc.), LLM framework (LangChain), and prompt engineering (Chain of Thought Prompting, Tree of Thoughts, Satisfiability-Aided Prompting).

Experience with explainable AI, responsible AI, data analysis, NLP, CV, and diverse libraries such as pandas, numpy, scikit-learn, TensorFlow, PyTorch, and PySAT.

Awards and Scholarships

2021-2024	**Monash Graduate Scholarship**	Scholarship covers living expenses and tuition
2020	**Best Paper Award**	Our paper "Computing Optimal Decision Sets with SAT" has been selected for the Best Paper Award for the CP/ML Track of CP 2020.

Scientific Activities

2024	PC member of the AAAI Conference on Artificial Intelligence(AAAI-2024)

References available upon request.