

자기소개서

전민석

March 25, 2024

모든 문제는 알맞은 언어를 디자인 및 사용하여 풀어야 한다. 문제와 알맞는 언어가 만나면, 문제의 본질을 꿰뚫어볼 수 있게 되고 그에 따라 문제를 제대로 해결하는 방법을 찾을 수 있게 되기 때문이다. 이는 컴퓨터학 분야에서도 마찬가지이다. 컴퓨터학 각 분야의 문제들의 본질을 꿰뚫어보고, 그에 따른 해결책을 찾기 위해 적합한 언어를 디자인하고 사용하는 것이 중요하다. 이 철학을 기반으로 컴퓨터학 분야의 다양한 문제들의 해결책을 표현하기 위한 언어들을 디자인해왔고, 언어 위에서 정답을 찾는 알고리즘을 개발해왔다. 디자인한 언어들은 문제의 본질적인 해결책을 만들 수 있도록 해주었고, 이를 통해 혁신적인 연구 및 수업 결과물들을 만들어 내었다. 컴퓨터학 분야에서 문제를 해결하기 위한 언어를 디자인한다는 것은 각 분야의 특화된 프로그래밍 언어(domain-specific programming language)를 디자인하는 것을 의미한다. 디자인 한 도메인 특화 프로그래밍 언어들은 문제의 정답을 프로그램으로써 표현할 수 있도록 디자인 되어야 하며, 이를 위해 문제의 본질적인 특성을 반영하여야 한다. 이러한 언어들은 각 문제의 정답(해결책)을 표현하기 위한 문법과 의미론을 제공하며, 디자인 한 언어 위에서 문제의 해결책을 찾아내는 알고리즘을 개발할 수 있게 해준다.

설명 가능한 그래프 기계학습, 소프트웨어 테스팅, 소프트웨어 정적 분석, 프로그래밍 교육 등 다양한 컴퓨터학 분야의 문제들의 본질적인 해결책을 제시하기 위해 도메인 특화 프로그래밍 언어들을 디자인하고 알고리즘을 개발해왔다. 설명 가능한 그래프 기계학습을 위해 디자인 한 특화 프로그래밍 언어는 본질적으로 설명 가능한 그래프 기계학습 모델을 표현할 수 있도록 디자인 되었으며, 이를 통해 학습한 그래프 기계학습 모델은 내놓은 모든 결과에 대해 올바른 설명을 제공할 수 있게 된다. 정적 분석을 위한 특화 프로그래밍 언어는, 분석을 빠르고 정확하게 해주는 정적 분석 전략을 표현할 수 있도록 디자인 되었으며, 소프트웨어 테스팅 문제를 해결하기 위한 특화 프로그래밍 언어는 효과적인 테스트 케이스를 쉽게 생성할 수 있도록 디자인 되었다. 학계에서는 디자인한 특화 프로그래밍 언어의 우수성을 인정 받아, 소위 최우수 국제 학술대회들(PLDI 2024, POPL 2022, OOPSLA 2020, OOPSLA 2018, OOPSLA 1017)에서 논문들을 발표해왔다. 제안한 언어 위에서 후속 연구들이 활발히 진행되고 있으며, 산업계에서도 제안한 언어를 사용하여 문제를 해결하고 있다. 프로그래밍 수업에서도 또한 수업에서 발생하는 문제들을 해결하기 위해 디자인한 특화 언어로 학생들이 프로그래밍을 배우면서 발생하는 문제들을 해결하고 있다.

1 설명 가능한 그래프 기계학습을 위한 특화 프로그래밍 언어

프로그램 분석, 이상 거래 탐지, 프로그램 수정, 약물 발견 등 다양한 현실 세계의 문제들은 그래프 문제로 표현될 수 있다. 이러한 문제들은 그래프 기계학습 문제로 모델링될 수 있으며, 현재 인공지능망 기반의 그래프 기계학습 모델(Graph Neural Networks)이 많은 그래프 문제들을 해결하고 있다 [1]. 그러나, 이러한 인공지능망 기반 모델들은 내놓은 결과에 대한 설명을 제공하지 않는 단점이 있다. 이상 거래 탐지 및 신약 개발과 같은 사람의 안전과 관련된 중요한 문제들에서는 모델의 결과에 대한 설명이 필수적이기 때문에, 인공지능망 기반 그래프 기계학습 방법은 적합하지 않다. 그래프 기계학습 모델의 결과에 대한 설명을 제공하는 방법들이 다양하게 제안되어 왔지만, 인공지능망 설명의 본질적 어려움 때문에 제안된 방법들 또한 모델의 결과에 대한 실제 이유를 반영하는 제대로 된 설명을 제공하지 못한다 [2].

Our approach can also be generalized to develop inherently explainable machine learning methods. Many significant real-world problems can be modeled as graph machine learning problems, including fraud detection, program repair, and drug discovery. In such decision-critical applications, it is important to understand why the models made each prediction. However, the dominant neural-network based graph machine learning models (i.e., graph neural networks) are black-box models that do not explain why they made each prediction. Various GNN explanation techniques have been proposed, but they fall short in providing satisfactory explanations because explaining the black-box models is fundamentally challenging. To address this problem, I developed a new inherently explainable machine learning method PL4XGL. The key idea of PL4XGL is to generate inherently explainable graph machine learning models by designing a domain-specific language, called GRAPH DESCRIPTION LANGUAGE, and synthesis algorithms that learn models from training graph data. The experimental results show that our DSL-based approach successfully generates inherently explainable graph machine learning models that show competitive accuracy compared to the existing dominant neural network-based models but provide significantly better explanations for the predictions. Based on this work, I will lead DSL-based explainable machine learning.

2 고성능 정적 분석을 위한 특화 프로그래밍 언어

3 고성능 소프트웨어 테스트를 위한 특화 프로그래밍 언어

4 프로그래밍 수업에서 발생하는 문제 해결을 위한 특화 프로그래밍 언어

5 최종 목표

References

- [1] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and Philip S. Yu. A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 32(1):4–24, 2021.
- [2] Hao Yuan, Haiyang Yu, Shurui Gui, and Shuiwang Ji. Explainability in graph neural networks: A taxonomic survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–19, 2022.