ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
UNIVERSITY OF PIRAEUS

ΤΜΗΜΑ ΨΗΦΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

Δ.Π.Μ.Σ. ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ

DEMOKRITOS
ΙΝΣΤΙΤΟΥΤΟ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

Προτεινόμενα Θέματα Διπλωματικής Εργασίας

Από

Νίκο Κατζούρη

A. Neuro-Symbolic Knowledge Learning

Neuro-symbolic (NeSy) AI aims at integrating neural learning from perceptual-level experience with reasoning through high-level knowledge. NeSy techniques can improve the robustness, generalisation capacity and interpretability of NNs, by encouraging their compliance with existing domain knowledge and constraints. NeSy systems are typically compositional, pairing a neural learning component with a symbolic, reasoning one. The former makes predictions from sub-symbolic input, while the latter reasons with the neural predictions through some background knowledge, to make high-level inferences. Consider for instance a vision model that extracts semantically meaningful features from images, coupled with a symbolic model that reasons with these features, in order to make high-level inferences for some downstream predictive task. Importantly, by making reasoning differentiable, the two components may be tightly integrated in a fashion that allows to train a NeSy system end-to-end, so that the neural predictions are aligned with the knowledge.

Most NeSy approaches assume that the symbolic part, i.e., the knowledge that is to be reasoned upon, is provided beforehand. In real-life applications, however, this is not always the case. Therefore, an important requirement in NeSy applications is the ability to extract high-level task semantics, in the form of symbolic knowledge, while training a network to make task-specific, low-level predictions. It is a challenging and under-explored problem, due to the fact that a "starting point" for the learning process is missing. A starting point could either be some concrete background knowledge, as in standard NeSy settings, or a pre-trained (or partially trained) network.

The purpose of this thesis is to explore techniques for mitigating difficulties in NeSy knowledge extraction via e.g. using NeSy training methods for simultaneously refining some crude, initial knowledge, while fine-tuning pre-trained networks w.r.t. the knowledge, such as foundation models, vision language models etc.

The project requires good knowledge of Python programming and a solid background on deep learning and knowledge representation & reasoning. The developed techniques may be applied on knowledge extraction for challenging applications, including autonomous driving, robot navigation and disease progression monitoring.

Bibliography:

Marra, G., Dumancic, S., Manhaeve, R. and De Raedt, L. From statistical relational to neurosymbolic artificial intelligence: A survey. Artificial Intelligence, 2024

Cunnington, D., Law, M., Lobo, J., & Russo, A., The Role of Foundation Models in Neuro-Symbolic Learning and Reasoning. arXiv preprint arXiv:2402.01889, 2024

Liu, A., Xu, H., Van den Broeck, G., & Liang, Y., Out-of-distribution generalization by neural-symbolic joint training, AAAI, 2023

B. **Neuro-Symbolic Techniques for Safe and Robust AI**

The deployment of AI systems in mission and safety-critical applications, such as autonomous driving and medical decision support, is predicated on techniques that can strengthen, or even guarantee their safety and robustness. Neuro-symbolic (NeSy) techniques can facilitate progress towards that direction, by encouraging the compliance of deep learning predictors (e.g. perception networks) with functional specifications of symbolic nature, related to safety and correctness. Consider for instance the ability to enforce symbolic safety constraints on the behaviour of a compositional autonomous system consisting of several neural components, related to perception, control, action selection etc.

The purpose of this thesis is to explore the application of existing approaches to safe and robust AI, such as safe sequential decision making, or NeSy verification, to challenging application domains of temporal nature, such as autonomous driving and robot navigation. To that end, existing techniques will be extended to a temporal setting where necessary, exploring also interactions with critical event detection and forecasting techniques.

The project requires good knowledge of Python programming and a solid background on deep learning and knowledge representation & reasoning.

Bibliography:

Yang, W. C., Marra, G., & De Raedt, L., Safe Reinforcement Learning via Probabilistic Logic Shields, IJCAI 2023

Giunchiglia, E., Tatomir, A., Stoian, M. C., & Lukasiewicz, T. CCN+: A neuro-symbolic framework for deep learning with requirements. *International Journal of Approximate Reasoning*, 2024.

Xie, X., Kersting, K., & Neider, D. (2022). Neuro-Symbolic Verification of Deep Neural Networks, IJCAI 2022

ΤΜΗΜΑ ΨΗΦΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΙΝΣΤΙΤΟΥΤΟ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ