ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
UNIVERSITY OF PIRAEUS

ΤΜΗΜΑ ΨΗΦΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

Δ.Π.Μ.Σ. ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ

DEMOKRITOS
ΙΝΣΤΙΤΟΥΤΟ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

Προτεινόμενα Θέματα Διπλωματικής Εργασίας

Από

Διδάσκοντα Χριστόφορο Ρεκατσίνα

A. Shallow machine learning architectures to enhance molecular simulations

Qualifications required: Python programming; Machine Learning

This thesis proposes a machine learning (ML) architecture designed to enhance and accelerate molecular simulations for nanoporous materials. The architecture integrates three core ML modules: a physics-guided neural network (PGNN), a physics-informed neural network (PINN), and a Gaussian process optimization algorithm. The PGNN operates at the molecular level, utilizing molecular coordinates as input and generating outputs that include the coordinates and the number of adsorbates. Its loss function is specifically designed to minimize the system's energy, guided by the fundamental physics equations that govern the molecular interactions. Building on this, the PINN scales the problem to the macro level, enabling applications to real-world scenarios. It achieves this by solving differential equations relevant to key processes such as adsorption and diffusion, thereby validating the results against experimental data. Concurrently, the Gaussian process optimization algorithm focuses on inverse design, aiming to optimize the material's performance at the molecular level by identifying configurations that maximize desired properties. Together, these modules form a comprehensive framework that bridges molecular-scale simulations with macroscopic applications, offering a robust and efficient approach to understanding and designing nanoporous materials.

Supervisors: Christoforos Rekatsinas, George Giannakopoulos, Panagiotis Krokidas

References

Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, https://doi.org/10.1016/j.jcp.2018.10.045

Inverse design of ZIFs through artificial intelligence methods, https://doi.org/10.1039/D4CP02488E

B. Advancing Bayesian Optimization: Efficient Strategies for Accelerated Materials Discovery

Qualifications required: Python programming; Machine Learning

Qualifications desired: Machine learning and Deep learning toolkits (e.g. PyTorch); genetic algorithms, bayesian optimization, active learning
Supervisors: George Giannakopoulos, Panagiotis Krokidas, Christoforos Rekatsinas
Contact: ggianna [at] iit.demokritos.gr, p.krokidas [at] iit.demokritos.gr
Related groups: INSANE Group

Description: Bayesian Optimization (BO) is an AI-driven technique designed to efficiently explore and sample large design spaces, particularly in domains where data are sparse, experiments are costly, and simulations are computationally intensive. This approach is especially valuable in the development of novel materials, enabling the discovery of optimized materials with enhanced performance while significantly reducing the number of physical or computational experiments required. However, traditional BO methodologies are constrained by high memory demands, largely due to the reliance on Gaussian Processes (GPs) as surrogate models. This limitation leads to a computational complexity that scales as $O(n^3)$, making it impractical for large-scale or high-dimensional problems. To address these challenges, next-generation BO algorithms must incorporate more efficient strategies. This thesis will focus on investigating advanced approaches to improve the efficiency of BO, such as batch sampling techniques, dynamically optimized batch sizes, adaptive exploration-exploitation trade-offs, efficient surrogate models, and multi-fidelity methods inspired by state-of-the-art advancements. The goal is to develop and evaluate innovative methods that maintain the predictive power of BO while significantly reducing memory and computational overhead, ultimately enabling its broader applicability in materials science and other high-impact fields.

References

Siemenn et al. Fast Bayesian optimization of Needle-in-a-Haystack problems using zooming memory-based initialization (ZoMBI), npj Computational Materials, 2023

Gantzler et al. Multi-fidelity Bayesian optimization of covalent organic frameworks for xenon/krypton separations, Digital Discovery, 2023

C. Predicting relevance in scientific information retrieval

Qualifications required: Python programming; Machine Learning and Deep Learning algorithms.

Qualifications desired: Machine learning and Deep learning toolkits (e.g. PyTorch); Natural Language Processing (NLP)
Supervisors: George Giannakopoulos, Artemis Dampa
Contact: ggianna [at] iit.demokritos.gr
Related groups: INSANE Group

Description: The objective of this thesis is to propose and implement methods to improve relevance prediction in scientific information retrieval. During scientific research and literature reviews, researchers often retrieve a large number of documents, datasets, and other resources that need to be managed in order to extract valuable, domain-specific knowledge. To address this challenge, this work will primarily focus on, but is not limited to, the effective representation of scientific documents. This can be achieved by incorporating prior knowledge to better capture the complex, specialized nature of scientific texts. Additionally, the thesis will explore unbiased methods for corpora collection, annotation, and relevance evaluation, as well as relevance classification techniques that integrate both explicit and implicit user feedback. Lastly, query expansion using natural language processing methods will be utilized to automatically expand search queries with related terms, synonyms, and conceptually relevant keywords. Ultimately, the aim is to develop a more accurate and efficient system for retrieving and evaluating scientific literature, enhancing researchers' ability to access the most relevant information.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
UNIVERSITY OF PIRAEUS

ΤΜΗΜΑ ΨΗΦΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

Π.Μ.Σ. ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ

DEMOKRITOS
ΙΝΣΤΙΤΟΥΤΟ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

D. How Explainable is Informed Machine Learning?

Qualifications required: Python programming; Machine Learning and Deep Learning algorithms

Supervisors: Christoforos Rekatsinas, Vasilis Gkatsis
Contact: crek [at] iit.demokritos.gr
Related groups: INSANE Group

Abstract
Modern Machine Learning (ML) algorithms require large amounts of training data in order to be efficient, so they struggle when applied in areas characterized by data scarcity. Training data may be limited due to the cost of creation, collection or annotation. In applied natural sciences for example acquiring new data may require performing complex experiments or simulations which can be very time consuming and even very costly. The field of Informed Machine Learning has been developed in order to cope with such situations. It consists of ML models which incorporate some part of problem related knowledge, which can vary from laws of nature, to experts' intuition. It has been found that embedding such knowledge in the ML pipeline results in better performing models, trained with less data, and more explainable than traditional ones. An ML model is considered explainable when the reasoning that led to the final decision can be completely or partially understood by humans. Informed ML explainability can be described as leaving the black box approaches (data-dependent) for more transparent versions (knowledge embedded). But is this explainability really useful? The task of this thesis is to study different methods of informed machine learning, study modern approaches on explainability evaluation and create relevant metrics in order to assess the explainability efficiency of different informed ML models as well as in contrast to traditional ones.

References

Laura von Rueden, Sebastian Mayer, Katharina Beckh, Bogdan Georgiev, Sven Giesselbach, Raoul Heese, Birgit Kirsch, Julius Pfrommer, Annika Pick, Rajkumar Ramamurthy, Michal Walczak, Jochen Garcke, Christian Bauckhage, and Jannis

Schuecker. Informed Machine Learning – A Taxonomy and Survey of Integrating Prior Knowledge into Learning Systems. IEEE Transactions on Knowledge and Data Engineering, 35(1):614–633, January 2023. Conference Name: IEEE Transactions on Knowledge and Data Engineering. Ribana Roscher, Bastian Bohn, Marco F. Duarte, and Jochen Garcke. Explainable Machine Learning for Scientific Insights and Discoveries. IEEE Access, 8:42200–42216, 2020.

Vilone, Giulia, and Luca Longo. "Notions of explainability and evaluation approaches for explainable artificial intelligence." *Information Fusion* 76 (2021): 89-106

E. Grammar-Based Generative Modeling for architected materials with Property Optimization

Qualifications required: Python programming; Machine Learning and Deep Learning algorithms

Supervisors: Christoforos Rekatsinas, Vasilis Vasilis Sioros, Panagiotis Krokidas
Contact: crek[at] iit.demokritos.gr
Related groups: INSANE Group

Abstract

Metal-Organic Frameworks (MOFs), and architected composite materials, hold significant promise for advanced industrial applications, including gas separation, H2 adsorption, enhanced ductility, and damping. Despite progress in predictive methods, challenges persist in designing both MOFs and architected composites with superior functional properties, such as high adsorption capacity, mechanical resilience, and energy dissipation. While deep learning (DL)-based generative models have shown potential for molecular and material design, their reliance on large datasets limits their applicability in data-scarce domains. Grammar-based generative approaches offer an interpretable and data-efficient alternative, enabling the explicit integration of chemical and structural constraints. This project aims to adapt a grammar-based generative method for the design of MOFs and architected composite materials, incorporating property prediction models to optimize structures for targeted applications, including gas separation, hydrogen storage, mechanical durability, and damping performance.

References

Data-Efficient Graph Grammar Learning for Molecular Generation

A comprehensive transformer-based approach for high-accuracy gas adsorption predictions in metal-organic frameworks

Bio-inspired discontinuous composite materials with a machine learning optimized architecture

F. Enhancing Mesh-Based Simulation with Graph Neural Networks: Integrating Physics for Improved Accuracy

Supervisors: Christoforos Rekatsinas, Ilias Zavitsanos, Dimitris Kelesis

This thesis investigates advanced mesh-based simulations using Graph Neural Networks (GNNs), inspired by the MeshGraphNets framework. The objective is to enhance the predictive accuracy of physical simulations by incorporating additional physics-informed features. The student will begin by reproducing the results of the reference paper using PyTorch, leveraging real-world data relevant to the problem domain. The next step involves developing an enhanced model that integrates more comprehensive physical information, with the goal of outperforming the baseline approach in terms of accuracy and generalization.

Qualifications required: Python programming (PyTorch, TorchGeometric), Machine Learning and Deep Learning algorithms

References

https://proceedings.neurips.cc/paper_files/paper/2023/file/70518ea42831f02afc3a2828993935ad-Paper-Conference.pdf

https://arxiv.org/pdf/2010.03409

ΤΜΗΜΑ ΨΗΦΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΙΝΣΤΙΤΟΥΤΟ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

G. Efficient Molecular Dynamics Sampling using Latent Diffusion

Supervisors: Christoforos Rekatsinas, Panagiotis Krokidas, Vaia Prassia
Contact: crek[at] iit.demokritos.gr
Related groups: INSANE Group

Abstract

Diffusion models (DMs) have demonstrated impressive performance in generating high-quality data by learning a probabilistic process that iteratively transforms noise into meaningful samples. In the realm of image generation, models like Latent Diffusion Models (LDMs) have achieved state-of-the-art results by operating in latent space, greatly reducing computational requirements while preserving detail. Inspired by this success, this project explores the application of Latent Diffusion to molecular dynamics (MD) simulations to improve the efficiency of biomolecular process simulations. DIFFMD, inspired by thermodynamic diffusion processes, provides a novel and computationally efficient method for simulating biomolecular dynamics without relying on traditional energy or force calculations. The goal of this thesis is to evaluate and enhance the performance of DIFFMD by applying latent diffusion techniques and comparing it to existing deep learning-based MD models. The work will involve integrating latent diffusion with a geometric Transformer and optimizing it for biomolecular applications. The project will also explore the potential for further scaling and improving the method to simulate a broader range of biomolecular processes.

References

High-Resolution Image Synthesis with Latent Diffusion Models

DiffMD: A Geometric Diffusion Model for Molecular Dynamics Simulations

H. Grounding First-Order Logic from Image Data for Human-Interpretable Knowledge Graph Construction

Abstract

This project aims to create a comprehensive knowledge database using First-Order Logic (FOL) extracted from image data. The primary contribution lies in learning predicates along with their arguments, enhancing the interpretability of the extracted knowledge as opposed to relying on anonymous, non-human interpretable symbols. The goal is to advance neurosymbolic AI by grounding FOL representations from raw visual inputs, providing more detailed and human-understandable representations of object interactions, relationships, and the environment. The project builds upon the First-Order State AutoEncoder (FOSAE) model, extending its capabilities to create scalable, semantically rich knowledge graphs that bridge neural perception and symbolic reasoning.

References

Unsupervised Grounding of Plannable First-Order Logic Representation from Images

TV-TREES: Multimodal Entailment Trees for Neuro-Symbolic Video Reasoning