ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
UNIVERSITY OF PIRAEUS
ΤΜΗΜΑ ΨΗΦΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

Δ.Π.Μ.Σ. ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ

DEMOKRITOS
ΙΝΣΤΙΤΟΥΤΟ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

Προτεινόμενα Θέματα Διπλωματικής Εργασίας

Από

Καθ. Γεώργιο Βούρο

A. Diffusion models for online model free reinforcement learning

This thesis aims to investigate the potential and advances of diffusion models for online model free reinforcement learning (RL), design and implement an RL method that exploits such models.

The stages are as follows:

- Study state of the art methods exploiting diffusion models in online model-free RL and beyond
- Experiment with the most promising methods, identify cons and prons
- Design and implement an online model-free RL method exploiting diffusion models.

References

[1] S.Zhu et al., Diffusion Models for Reinforcement Learning: A Survey, 2024, https://arxiv.org/abs/2311.01223

[2] H.Ma et al., Soft Diffusion Actor-Critic: Efficient Online Reinforcement Learning for Diffusion Policy, 2025, https://arxiv.org/abs/2502.00361

[3] Y.Dong et al., Enhancing Sample Efficiency in Online Reinforcement Learning via Policy-Guided Diffusion Models, 2024, https://dl.acm.org/doi/full/10.1145/3704323.3704390

B. Learning to collaborate

This thesis aims to answer the following question: Given demonstrations from a potential collaborator, can an agent decide the best policy to be used jointly with the other, so as their joint action to result to stable and mutually-beneficial action?

References

[1] N.Koliou, G.Vouros, Ranking Joint Policies in Dynamic Games using Evolutionary Dynamics, AAMAS 2025

[2] S. Omidshafiei et al., α-Rank: Multi-Agent Evaluation by Evolution, Nature 2019, https://www.nature.com/articles/s41598-019-45619-9

[3] Y. Du, Estimating ↵-Rank from A Few Entries with Low Rank Matrix Completion, 2021 https://discovery.ucl.ac.uk/id/eprint/10174057/

C. Modelling for human-AI collaboration.

This thesis aims to investigate answering the question how can we build reinforcement learning algorithms that optimize jointly own utility, reflecting their objective, as well the utility of humans? This need is emergent in safety-critical systems, where humans need to trust the agent and the agent needs to learn human preferences and needs and align with them.

There are different ways to try to answer this question. The thesis will investigate alternatives based on reinforcement learning with human feedback (e.g. [1]), models calibration (e.g. [2]) and conformal prediction (e.g. [3]).

References

[1] D.Ziegler et al., "Fine-Tuning Language Models from Human Preferences", 2019

[2] K Vodrahalli et al., "Uncalibrated Models Can Improve Human-AI Collaboration", 2022

[3] A. Angelopoulos and S.Bates, A Gentle Introduction to Conformal Prediction and Distribution-Free Uncertainty Quantification, 2021.

D. Offline Safe Reinforcement Learning with the Decision Transform.

This thesis aims to study state of the art on the use of Decision Transformer for Safe Reinforcement Learning and go beyond that.

References

[1] Lili Chen et al., 2021, Decision Transformer: Reinforcement Learning via Sequence Modeling, https://arxiv.org/abs/2106.01345

[2] Zuxin Liu, et al., 2023, Constrained Decision Tranformer for Offline Safe Reinforcement Learning, https://proceedings.mlr.press/v202/liu23m.html

E. Revisiting Behavioral Cloning.

ΤΜΗΜΑ ΨΗΦΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΙΝΣΤΙΤΟΥΤΟ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

This thesis aims to revisit the behavioural cloning technique taking advantage of latest achievement in machine learning and prediction, addressing distribution shift, compounding errors, causal confusion and need of abundance of data.

References

[1] Ross et al 2011, A Reduction of Imitation Learning and Structured Prediction to No-Regret Online Learning

[2] De Haan et al, 2019, Causal confusion in imitation learning

[3] Shah at al, 2023, GNM : A general Navigation model to drive any robot