ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
UNIVERSITY OF PIRAEUS

DEMOKRITOS

Προτεινόμενα Θέματα Διπλωματικής Εργασίας

Από

Καθηγητή Γεώργιο Βούρο

A. Explainability using Decision Trees and Stochastic Gradient Trees

Abstract:

Stochastic Gradient Trees[8] have shown efficacy in representing the policy of an agent, while offering interpretability in decision making.

The purpose of this thesis is to build a toolbox for extracting explanations for local agent decisions and for the global agent function (policy) from decision trees, with special emphasis on Stochastic Gradient Trees.

The toolbox must at least

1. Extract arguments for/against local decisions
2. Provide the importance of features for local decisions
3. Provide an overview of arguments for/against decisions

using domain-independent methods.

[8] Henry Gouk, Bernhard Pfahringer, Eibe Frank, Stochastic Gradient Trees, https://arxiv.org/abs/1901.07777

B. Explainable Deep Reinforcement Learning with Interpretable Models

Abstract:

While deep reinforcement learning (DRL) methods incorporate deep models that do not offer inherent interpretability, such models can be replaced by inherently interpretable models (e.g. decision trees) offering solutions to the overall interpretability of DRL methods[9]. The so said trade-off between efficacy and interpretability of models has been challenged by many works that argue that DRL methods with interpretable models can be as effective as DRL methods with deep, non-interpretable models.

This thesis aims to address and provide evidence that indeed, we can construct inherently interpretable reinforcement learning methods, that are as effective as the deep, non-interpretable ones.

Γ. Solving Regression Problems using deep convolutional neural networks

Abstract:

This thesis aims at studying deep convolutional neural network (DCNN) towards solving important problems in the estimation of hidden variables' regarding the evolution of time series (e.g. of trajectories) using regression methods.

In such problems, the history of observations is important and predictions are taken by exploiting such histories.

This thesis will explore the use of DCNN algorithms for regression, using real-world data sets in important domains (e.g. aviation).

[9] G. Vouros, "Explainable Deep Reinforcement Learning: State of the Art and Challenges", https://dl.acm.org/doi/abs/10.1145/3527448