Deep Reinforcement Learning for Data Management in Hierarchical Storage System

Tianru Zhang, Salman Toor *

Department of Information Technology, Uppsala university

Project Description

Large-scale data management is one of the most challenging prospects in the big data domain [1]. Over the years, a number of frameworks for data management have become available. Most of them are effective, but with weak flexibility. It becomes difficult to move and work coherently with data as new requirements emerge. A possible solution is to use an intelligent hierarchical (multi-tier) storage system (HSS). A HSS is a meta solution that consists of different storage mediums organized as a jointly constructed storage pool. The main idea is to connect different independent storage solutions and move the data between them according to a data migration policy designed to meet a set of requirements. In one of our recent articles [2], we have comprehensively described the underlying challenges related to storage hierarchies and introduced an open-source hierarchical storage framework with a dynamic migration policy based on reinforcement learning (HSM-RL). Our experiments results based on static and dynamic datasets and the results have proved the effectiveness, efficiency and consistency of the RL-based policy. However, given special use cases such as scientific imaging experiments where internal features of files are difficult to obtain, previous defined RL-based policy that heavily used file features become challenging. Therefore, we consider the Deep RL, where deep neural networks are used to solve the problem of value function approximation given indeterminable states.

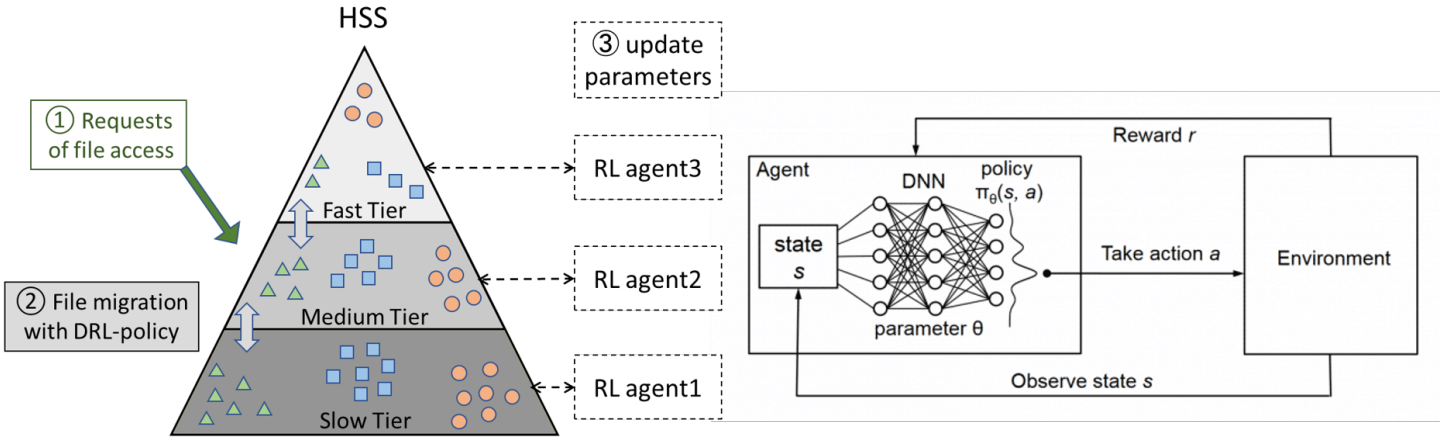

Figure1: An example of three tiers hierarchical storage system (HSS), with DRL-based data migration policy. On the right side is the Deep RL agents where deep Q network is used to approximate the value function

In this project, the aim is to first understand the hierarchical storage management (HSM) problem and deep reinforcement learning (DRL). Then it is expected to apply DRL methods such as Deep Q Network to solve the HSM problem. Finally the work of this project will continue to be a research publication. Good programming skills in Python are necessary, and knowledge about reinforcement learning is preferred. Preliminary courses related to numerical linear algebra and machine learning are also required.

References

[1] Vivien Marx. "The big challenges of big data". In: *Nature* 498.7453 (2013), pp. 255–260. DOI: https://doi.org/10.1038/498255a (cit. on p. 1).

[2] Tianru Zhang, Andreas Hellander, and Salman Toor. "Efficient Hierarchical Storage Management Empowered by Reinforcement Learning". In: *IEEE Transactions on Knowledge and Data Engineering* (2022), pp. 1–1. DOI: 10.1109/TKDE.2022.3176753 (cit. on p. 1).

*tianru.zhang@it.uu.se, salman.toor@it.uu.se