Cache Replacement Using Reinforcement learning

Team Members: Rohit Kumar (ZR66734)
Arti Singh (LN22810)
Aishwarya Sapkale (DP73694)

Cache Replacement

- A program/utility to manage stored information on cache
- Replacement Policies:
  - *Random Replacement*
  - *LRU (Least Recently Used) / TLRU (Time aware LRU)*
  - *First In First Out*
  - *MRU (Most Recently Used)*
  - *Least-frequently Used*
  - *Bélády's Algorithm*

Bélády's Algorithm

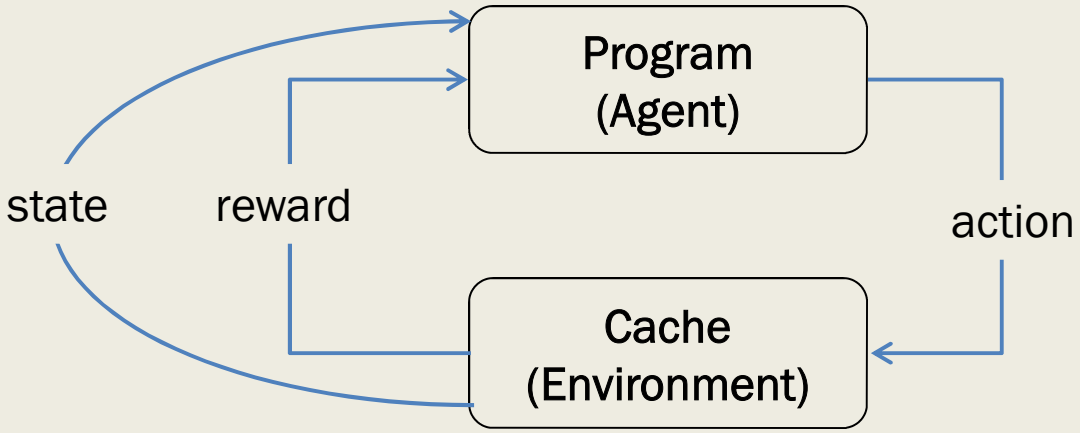
- Discards the information not to be used in near future.
  - *Information with low probability of being used.*
- Practically possible with use of Machine Learning
- Deep learning has been applied recently for hardware predictors.
- E.g. RNN, Multi-layer Perceptron, CNN

Motivation

- Are earlier algorithms optimal for replacement?
  - *Past is not always reliable*
- Is predicting future feasible for optimizing the replacement policies?
  - *Yes, using predictive modeling*
  - *However, it takes enormous resources to train*
  - *Training can take up to hours, days, or even months.*
  - *Also, the models have fixed learnt parameters, so not adaptable*
  - *State of cache is never constant and is changing over time.*

Proposed Idea

- Applying Reinforcement Learning for Cache Replacement using Q-learning algorithm



Methodology

- Goal – Reduce the cache miss rate.
- For k initial time-steps
  - *select random cache entries to be replaced*
  - *keep a record of each time-step*
- Calculate the overall cache miss rate.
  - *If it is below the threshold -> Add reward*
  - *Otherwise -> Reduce reward*
  - *If same value as threshold -> No change in reward*
- Update the Q-table based on the reward.
- Iterate over above steps, based on Q-table entries instead now for selection of entry to be replaced

References

- Deep Reinforcement Learning for Adaptive Caching in Hierarchical Content Delivery Networks [July 2019, Alireza Sadeghi, Gang Wang, and Georgios B. Giannakis https://arxiv.org/pdf/1902.10301.pdf
- A Q-learning-based network content caching method [Dec 2018: Haijun Chen and Guanzheng Tan] https://www.researchgate.net/publication/329219973_A_Q-learning-based_network_content_caching_method
- Applying Deep Learning to the Cache Replacement Problem [Oct 2019: Zhan Shi, Xiangru Huang, Akanksha Jain, and Calvin Lin] https://www.cs.utexas.edu/~lin/papers/micro19c.pdf
- https://wiki.ubc.ca/Better_caching_using_reinforcement_learning

Thank You