AI-based Cache Replacement in Virtualized Environment

Taisei MIURA[†a)], ***Non member***, **Kenichi KOURAI**[††], ***and*** **Saneyasu YAMAGUCHI**[†], ***Member***

1. Introduction

Deep learning has been producing breakthroughs in many fields, especially in image processing and natural language processing. However, it has been used in a restrained manner in a system software field. For example, deep learning may be suitably supported in cache replacement, but replacement using this has not been well studied.

In usual cases, the least recently used (LRU) algorithm, which is traditionally and widely used, works very effectively. However, in the host operating system cache in a virtualized environment, LRU does not work well [1]. In particular, the performance of LRU is severely low if the guest cache size is larger than the host cache size. Therefore, a proposal of a new algorithm, for example, an algorithm based on deep learning for such a situation is expected.

In this paper, we focus on a cache replacement based on deep learning and discuss its performance.

2. Related Work

Shi et al. discussed a cache replacement method based on deep learning [2]. They improved the accuracy of Hawkeye [3]. They then showed that their method outperformed the state-of-the-art. However, they did not discuss performance in a virtualized environment.

In the work of [1], we evaluated the performance of LRU in a host cache and mentioned the potential performance of deep learning in a host cache. In the work of [1], the performance was evaluated only with random accesses based on Zipf distribution. On contrary, we evaluated it with a practical access log in this paper.

3. Performance Evaluation

We evaluated the cache hit ratio in a host cache under a guest cache in a two-level cache hierarchy with a simulation. In the simulation, an application issued I/O requests according to an I/O trace log of a practical system.

Fig. 1 shows the simulation results. These indicate that the replacement according to deep learning outperformed LRU in situations wherein the guest cache size was larger than the host cache size. Namely, in the cases the guest cache size was 40 blocks, which are drawn with bold lines, the long short-term memory (LSTM) method outperformed LRU if the host cache size was 20, 30, or 40.

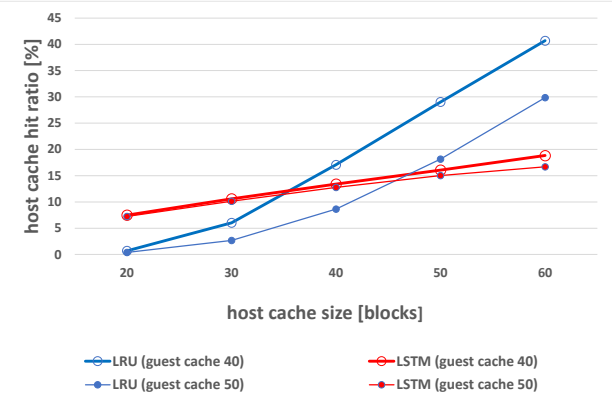

Fig. 1 Host cache hit ratio (LRU and LSTM)

4. Conclusion

In this paper, we discussed applying deep learning to cache replacement in a virtualized environment. We then showed that this method can achieve better performance than LRU in some cases.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Numbers 18K11277, 21K11854.

References

[1] T. Miura, K. Kourai and S. Yamaguchi, "Cache Replacement Based on LSTM in the Second Cache in Virtualized Environment," 2020 Eighth International Symposium on Computing and Networking Workshops (CANDARW), 2020, pp. 421-424, doi: 10.1109/CANDARW51189.2020.00086.

[2] Zhan Shi, Xiangru Huang, Akanksha Jain, and Calvin Lin. 2019. Applying Deep Learning to the Cache Replacement Problem. In Proceedings of the 52nd Annual IEEE/ACM International Symposium on Microarchitecture (MICRO '52). Association for Computing Machinery, New York, NY, USA, 413–425. DOI:https://doi.org/10.1145/3352460.3358319

[3] Akanksha Jain and Calvin Lin. 2016. Back to the future: leveraging Belady's algorithm for improved cache replacement. In 43nd Annual International Symposium on Computer Architecture (ISCA). IEEE, 78ś89.

[†]The author is with Kogakuin University …

[††] The author is with Kyushu Institute of Technology ...

[a)] E-mail: cm20048@ns.kogakuin.ac.jp