

## REFERENCES

- [1] R. F. Olanrewaju, A. Baba, B. U. I. Khan, M. Yaacob, A. W. Azman, and M. S. Mir, "A study on performance evaluation of conventional cache replacement algorithms: A review." Institute of Electrical and Electronics Engineers Inc., 2016, pp. 550–556.
- [2] H. Choi and S. Park, "Learning future reference patterns for efficient cache replacement decisions," *IEEE Access*, vol. 10, pp. 25 922–25 934, 2022.
- [3] J. Jose and N. Ramasubramanian, "Applying machine learning to enhance the cache performance using reuse distance," *Evolutionary Intelligence*, vol. 16, pp. 1195–1216, 8 2023, this paper talks about the requirement to have large dataset when it comes to deep learning cache replacement policies comparing to other methods. Refer the highlighted sections.
- [4] V.-N. Pham, M. Josh, D.-T. Le, S. W. Lee, and H. Choo, *A Prediction-Based Cache Replacement Policy for Flash Storage*, 11 2021, pp. 161–169.
- [5] M. Yang, C. Yang, and J. Shao, *Transformer-Based Cache Replacement Policy Learning*, 11 2022, pp. 493–500.
- [6] M. Murugesan and E. Kirubakaran, "Impact of machine learning techniques with cache replacement algorithms in enhancing the performance of the web server," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 7, pp. 812–816, 6 2017. [Online]. Available: [http://ijarcsse.com/docs/papers/Volume\\_7/6\\_June2017/V7I6-0323.pdf](http://ijarcsse.com/docs/papers/Volume_7/6_June2017/V7I6-0323.pdf)
- [7] J. Shuja, K. Bilal, W. Alasmay, H. Sinky, and E. Alanazi, "Applying machine learning techniques for caching in next-generation edge networks: A comprehensive survey," *Journal of Network and Computer Applications*, vol. 181, p. 103005, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1084804521000321>
- [8] L. Abdo, I. Ahmad, and S. Abed, "A smart admission control and cache replacement approach in content delivery networks," 08 2022.
- [9] Y. Wang, Y. Yang, and Q. Wang, "An efficient intelligent cache replacement policy suitable for pacs," *International Journal of Machine Learning and Computing*, vol. 11, pp. 250–255, 05 2021.
- [10] P. P and R. A, "Machine learning-based cache replacement policies: A survey," *International Journal of Engineering and Advanced Technology*, vol. 10, pp. 19–22, 08 2021.

- [11] L. V. Rodriguez, F. Yusuf, S. Lyons, E. Paz, R. Rangaswami, J. Liu, M. Zhao, and G. Narasimhan, *Learning Cache Replacement with CACHEUS*. [Online]. Available: [www.usenix.org/conference/fast21/presentation/rodriguez](http://www.usenix.org/conference/fast21/presentation/rodriguez)
- [12] A. Jain and C. Lin, “Hawkeye: Leveraging belady’s algorithm for improved cache replacement.”
- [13] Z. Shi, X. Huang, A. Jain, and C. Lin, “Applying deep learning to the cache replacement problem,” in *Proceedings of the 52nd Annual IEEE/ACM International Symposium on Microarchitecture*, ser. MICRO ’52. New York, NY, USA: Association for Computing Machinery, 2019, p. 413–425. [Online]. Available: <https://doi.org/10.1145/3352460.3358319>
- [14] E. Liu, M. Hashemi, K. Swersky, P. Ranganathan, and J. Ahn, “An imitation learning approach for cache replacement,” in *Proceedings of the 37th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, H. D. III and A. Singh, Eds., vol. 119. PMLR, 13–18 Jul 2020, pp. 6237–6247. [Online]. Available: <https://proceedings.mlr.press/v119/liu20f.html>
- [15] S. Sethumurugan, J. Yin, and J. Sartori, “Designing a cost-effective cache replacement policy using machine learning,” 02 2021, pp. 291–303.
- [16] H. J. Yoo, J. H. Kim, and T. H. Han, “RL-based cache replacement: A modern interpretation of belady’s algorithm with bypass mechanism and access type analysis,” *IEEE Access*, vol. 11, pp. 145 238–145 253, 2023.
- [17] Y. Wang, “Application of machine learning algorithm in cache replacement strategy,” in *2021 12th International Symposium on Parallel Architectures, Algorithms and Programming (PAAP)*, 2021, pp. 183–187.
- [18] G. Vietri, L. V. Rodriguez, W. A. Martinez, S. Lyons, J. Liu, R. Rangaswami, M. Zhao, and G. Narasimhan, “Driving cache replacement with ML-based LeCaR,” in *10th USENIX Workshop on Hot Topics in Storage and File Systems (HotStorage 18)*. Boston, MA: USENIX Association, Jul. 2018.
- [19] N. Megiddo and D. Modha, “Outperforming lru with an adaptive replacement cache algorithm,” *Computer*, vol. 37, pp. 58 – 65, 05 2004.
- [20] M. Zahran, “Cache replacement policy revisited,” 01 2007.
- [21] J. M. Kim, J. Choi, J. Kim, S. H. Noh, S. L. Min, Y. Cho, and C. S. Kim, “A Low-Overhead, High-Performance unified buffer management scheme that exploits sequential and looping references,” in *Fourth Symposium on Operating Systems Design and Implementation (OSDI 2000)*. San Diego, CA: USENIX Association,

- Oct. 2000. [Online]. Available: <https://www.usenix.org/conference/osdi-2000/low-overhead-high-performance-unified-buffer-management-scheme-exploits>
- [22] W. Ali and S. M. Shamsuddin, “Neuro-fuzzy system in partitioned client-side web cache,” *Expert Syst. Appl.*, vol. 38, pp. 14 715–14 725, 11 2011.
- [23] P. Vila, P. Ganty, M. Guarnieri, and B. Köpf, “Cachequery: learning replacement policies from hardware caches,” in *Proceedings of the 41st ACM SIGPLAN Conference on Programming Language Design and Implementation*, ser. PLDI 2020. New York, NY, USA: Association for Computing Machinery, 2020, p. 519–532. [Online]. Available: <https://doi.org/10.1145/3385412.3386008>
- [24] J. Yang, Y. Yue, and K. V. Rashmi, “A large scale analysis of hundreds of in-memory cache clusters at twitter,” in *14th USENIX Symposium on Operating Systems Design and Implementation (OSDI 20)*. USENIX Association, Nov. 2020, pp. 191–208. [Online]. Available: <https://www.usenix.org/conference/osdi20/presentation/yang>
- [25] B. Reed and D. D. E. Long, “Analysis of caching algorithms for distributed file systems,” *SIGOPS Oper. Syst. Rev.*, vol. 30, no. 3, p. 12–21, jul 1996. [Online]. Available: <https://doi.org/10.1145/230908.230913>
- [26] W. Ahmed, A. Shah, and P. Islamabad, “Cache memory: An analysis on optimization techniques,” *International Journal of Computer and Information Technology(2279-0764)*, vol. 04, pp. 414–418, 03 2015.
- [27] S. Podlipnig and L. Böszörmenyi, “A survey of web cache replacement strategies,” *ACM Comput. Surv.*, vol. 35, pp. 374–398, 12 2003.
- [28] M. Zulfa, A. Fadli, A. Permanasari, and W. Ali, “Performance comparison of cache replacement algorithms onvarious internet traffic,” *JURNAL INFOTEL*, vol. 15, pp. 1–7, 02 2023.
- [29] R. yan Cai, Y. Qian, and D. bin Wei, “Dynamic cache replacement strategy of space information network based on cache value,” *Journal of Physics: Conference Series*, vol. 2290, no. 1, p. 012106, jun 2022. [Online]. Available: <https://dx.doi.org/10.1088/1742-6596/2290/1/012106>
- [30] Y. Zha, P. Cui, Y. Hu, L. Xue, J. Lan, and Y. Wang, “An ndn cache-optimization strategy based on dynamic popularity and replacement value,” *Electronics*, 2022. [Online]. Available: <https://api.semanticscholar.org/CorpusID:252516306>
- [31] P. N. V. Kumar and V. R. Reddy, “Impact factor: 1.852 international journal of engineering sciences research technology novel web proxy cache replacement

- algorithms using machine learning techniques for performance enhancement,” *C)International Journal of Engineering Sciences Research Technology*, vol. 3, pp. 339–346, 2014. [Online]. Available: <http://www.ijesrt.com>
- [32] H. Sun, Q. Cui, J. Huang, and X. Qin, “Ncache: A machine-learning cache management scheme for computational ssds,” *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 42, no. 6, pp. 1810–1823, 2023.
- [33] Y. P. Pujari, M. Shaikshavali, L. N. Velayduam, V. Reddy, S. Kumar, and N. Kumar, “Selection of cache replacement algorithm for postgresql,” 03 2023.
- [34] Cache2k. (2022) Benchmarks for java in memory caches. [Online]. Available: <https://github.com/cache2k/cache2k-benchmark>
- [35] R. Chandra, M. Frean, and M. Zhang, “Adapting modularity during learning in cooperative co-evolutionary recurrent neural networks,” *Soft Computing-A Fusion of Foundations, Methodologies and Applications*, pp. 1–12, 06 2012.
- [36] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [37] sklearn. (2024) Incremental pca. [Online]. Available: <https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.IncrementalPCA.html>
- [38] K. W. CHURCH, “Word2vec,” *Natural Language Engineering*, vol. 23, no. 1, p. 155–162, 2017.
- [39] J. Pennington, R. Socher, and C. D. Manning, “Glove: Global vectors for word representation,” in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532–1543.
- [40] T. Hirasawa, M. Kaneko, A. Imankulova, and M. Komachi, “Pre-trained word embedding and language model improve multimodal machine translation: A case study in multi30k,” *IEEE Access*, vol. 10, pp. 1–1, 01 2022.
- [41] M. R. Costa-Jussà, Álvaro Nuez, and C. Segura, “Experimental research on encoder-decoder architectures with attention for chatbots,” *Computacion y Sistemas*, vol. 22, pp. 1233–1239, 2018.
- [42] S. Kandambige. (2024) Advanced cache replacement. [Online]. Available: <https://github.com/supun-kanda/advanced-cache-replacement>
- [43] ——. (2024) Msc project notebooks. [Online]. Available: <https://drive.google.com/drive/folders/1ZQmY87rtQrjssN8J1oeyKrW3eFELFmp9>

- [44] TensorFlow. (2024) Google cloud gpu pricing. [Online]. Available: <https://cloud.google.com/compute/gpus-pricing>
- [45] Keras. (2024) Adam optimizer. [Online]. Available: <https://keras.io/api/optimizers/adam>
- [46] ——. (2024) Probabilistic losses: Sparsecategoricalcrossentropy class. [Online]. Available: [https://keras.io/api/losses/probabilistic\\_losses/#sparsecategoricalcrossentropy-class](https://keras.io/api/losses/probabilistic_losses/#sparsecategoricalcrossentropy-class)