IMPROVING CACHE REPLACEMENT WITH MACHINE LEARNING

K.S Devinda

229314J

Masters in Computer Science

Department of Computer Science
Faculty of Engineering

University of Moratuwa
Sri Lanka

June 2024

IMPROVING CACHE REPLACEMENT WITH MACHINE LEARNING

K.S Devinda

229314J

Dissertation submitted in partial fulfillment of the requirements for the degree
Masters in Computer Science

Department of Computer Science
Faculty of Engineering

University of Moratuwa
Sri Lanka

June 2024

DECLARATION

I declare that this is my own work and this Dissertation does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or Institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature *UOM Verified Signature* Date: 2024-06-13

The supervisor should certify the Dissertation with the following declaration.

The above candidate has carried out research for the Masters in Computer Science Dissertation under my supervision. I confirm that the declaration made above by the student is true and correct.

Name of Supervisor: Prof. G.I.U.S. Perera

Signature of the Supervisor: Date: 14.06.2024

DEDICATION

I dedicate this research to my pillar of success, my wife Ojithma, who has been my strength and supported me throughout this course, enabling me to complete it.

ACKNOWLEDGEMENT

I dedicate this research to my supervisor, Prof. Indika Perera, for the guidance and valuable feedback provided, which enabled me to transcend this project.

ABSTRACT

The performance of caching systems heavily relies on the effectiveness of cache replacement policies, which aim to optimize the usage of cache storage and improve cache hit rates. Traditional cache replacement policies such as Least Recently Used (LRU) and Least Frequently Used (LFU) have been widely used but have limitations in adapting to changes in the workload and minimizing the use of expensive storage. Recently, researchers have proposed machine learning-based approaches to cache replacement, aiming to learn from the access history of cache blocks and make accurate predictions about the usefulness of cache blocks. These approaches have shown promising results in improving the cache hit rate and reducing the overall storage cost compared to traditional policies. The proposed solutions include various machine learning algorithms such as Linear Regression (LR), K-Nearest Neighbor (KNN), and Deep Learning (DL).

The proposed solution implements a dataset-specific prediction model which first derives a dataset using an algorithm which uses future references with very high accuracies. Then the supervised machine learning model will be used to derive the patterns on the derived dataset which has used the future reference pattern in cache replacement decision making process. By this method, the cache replacement strategy will be much more resilient as all the traditional and state of the art models uses cache data from the past only. The evaluation and the future work provides much stronger impression on this method as there are several factors of this research which may significantly improve the results and will help the current cache replacement strategies tremendously.

The document also explains about the technical environments which used to evaluate models. The machine learning model training and the dataset generation require significant memory and CPU performance requirement. Hence, dedicated GPU is required in some scenarios which this document explains on how to manage.

Keywords: Cache replacement, machine learning, performance optimization, cache

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES