An Optimization technology of Software Auto-Tuning Applied to Machine Learning Software

Toshiki Tabeta
Kogakuin University
Tokyo, Japan

Naoto Seki
Akihiro Fujii
Teruo Tanaka
Kogakuin University
Tokyo, Japan

Hiroyuki Takizawa
Tohoku University
Miyagi, Japan

1 INTRODUCTION

This study aims to enhance the performance of an application program by software auto-tuning. Performance factors of a target program are formed into multiple performance parameters. In software auto-tuning, it requires a long duration to explore combinational patterns of performance parameters [1]. We have proposed "DSICE", the D-Spline within Iterative Collinear Exploration, to operate the technique of performance parameter tuning. In our previous studies, the tuning techniques to estimate a combination of multiple performance parameters are built in ppOpen-AT. We develop a simple auto-tuning tool "DSICE", built independently from ppOpen-AT. DSICE allows to converge in a shorter duration even if combinational patterns increase exponentially, and estimate an appropriate combination [2]. In this study, we implement and apply DSICE to estimating hyperparameters' configuration of a machine learning model. The results of our experiment show DSICE's efficiency for optimization of a target program.

2 THE AUTO-TUNING TOOL "DSICE"

One-dimensional d-Spline is a discrete approximate function, and follows some collinear values sampled from total N to estimate the optimum value for a performance parameter. DSICE iterates directional exploration and collinear exploration using one-dimensional d-Spline. DSICE can take an advantage that memory consumption and computational complexity for one-dimensional d-Spline increase at a rate of $O(N)$ depending on individual performance parameters, not their combination [3]. Fig.1 shows the procedure of DSICE. DSICE automatically sets a promising combination of performance parameters for a target program, executes the program, and evaluates the performance in terms of a given metric. These steps are iterated until an acceptable combination is found.

3 THE APPLICATION TO MACHINE LEARNING

We have applied DSICE to tuning the hyperparameters ' configuration of a CNN model for an object classification problem called CIFAR10 reported in [4]. The evaluation model and parameters are based on those in [4]. In this study, DSICE optimizes the metric considering both of classification accuracy and learning time. The hyperparameters of CNN tuned in this study are summarized in Table 1. The total number of configuration patterns is 164,025. The results of our experiment are shown on Fig.2 where the vertical axis indicates the metric value minimized by DSICE. It is demonstrated that DSICE converges within about 130 iterations to sample metric value, and estimates an appropriate configuration that reduces metric value to about 0.3.

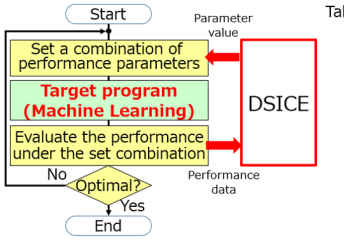

Fig.1 Procedure of DSICE

Tabel1 Hyperparameters of CNN

Parameters	Patterns
Filters (1st conv2d layer)	5
Output channels (1st conv2d layer)	9
Filters (2nd conv2d layer)	5
Output channels (2nd conv2d layer)	9
Neurons (3rd layer)	9
Neurons (4th layer)	9

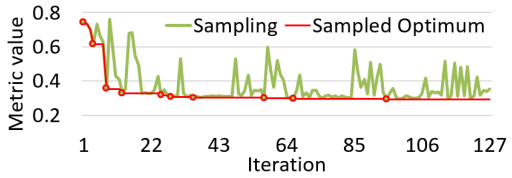

Fig.2 Transition of estimated optimum value

4 CONCLUSION

We have implemented an auto-tuning tool DSICE to estimate an appropriate combination of multiple performance parameters for a target program. As for the estimation of hyperparameters' configuration of machine learning consisting of 164,025 patterns, DSICE converges within 130 iterations and estimated an appropriate configuration. We will apply DSICE to other performance optimization problems in our future work.

REFERENCE

[1] T. Katagiri, Auto-tuning for The Era of Relatively High Bandwidth Memory Architectures: A Discussion Based on an FDM Application, Proceedings of IEEE IPDPSW2018.
[2] M. Mochizuki, A. Fujii, T. Tanaka, "Fast Multidimensional Performance Parameter Estimation with Multiple One dimensional d-Spline Parameter Search", The 12th Int'r Workshop on Automatic Performance Tuning (iWAPT2017).
[3] T. Tanaka, R. Otsuka, A. Fujii, T. Katagiri, T. Imamura, "Implementation of d-Spline-based incremental performance parameter estimation method with ppOpen-AT", Scientific Programming, Vol.22, No.4, pp.299-307, 2014.
[4] Z. Wang, M. Agung, R. Egawa, R. Suda, H. Takizawa, "Automatic Hyperparameter Tuning of Machine Learning Models under Time Constraints", The 2nd Int'r Workshop on Automation in Machine Learning and Big Data (AutoML 2018).