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

Toshiki Tabeta†, Naoto Seki†, Akihiro Fujii†, Teruo Tanaka†, Hiroyuki Takizawa‡
†Kogakuin University, ‡Tohoku University

HPC Asia 2020

TOHOKU UNIVERSITY

KOGAKUIN UNIVERSITY

Introduction

This poster shows to enhance the performance of an application program by software auto-tuning (AT). Performance factors of a target program are formed into performance parameters. So that to enhance performance is said that to estimate an appropriate combination of performance parameters. The estimation while executing a target program should be designed, to set a combination before each execution and evaluate the performance after each execution. For the estimation, we develop a simple AT tool "DSICE". The estimation by DSICE is based on the method of iterative collinear exploration using d-Spline [1][2]. The estimation method is built in the AT infrastructure ppOpen-AT [3]. DSICE makes the estimation method to be used in a more general form. Then we apply DSICE to estimating hyperparameters' configuration of a machine learning model.

The AT Tool "DSICE"

DSICE : D-Spline within Iterative Collinear Exploration

- DSICE as a function embedded into a target program

- Input
  - Combinational patterns of performance parameters
  - A single metric value (the lower the better) to evaluate the performance

- Output
  - An optimum combination under the input metric among the input patterns

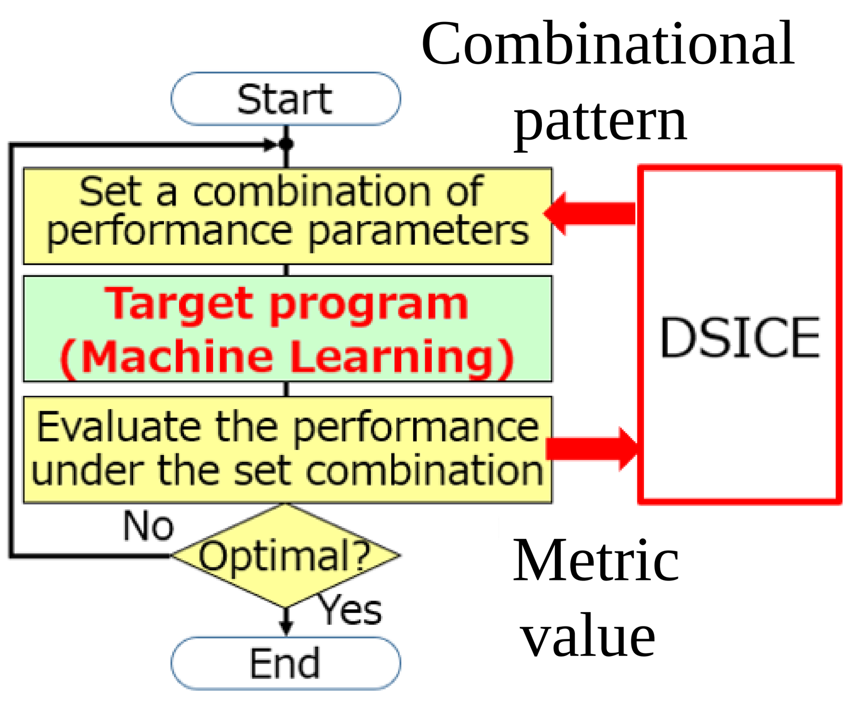

Fig.1 Procedure of DSICE: that iterates determining a combinational pattern of performance parameters among the input patterns, and sampling a metric value under the set combination, until an acceptable combination is found.

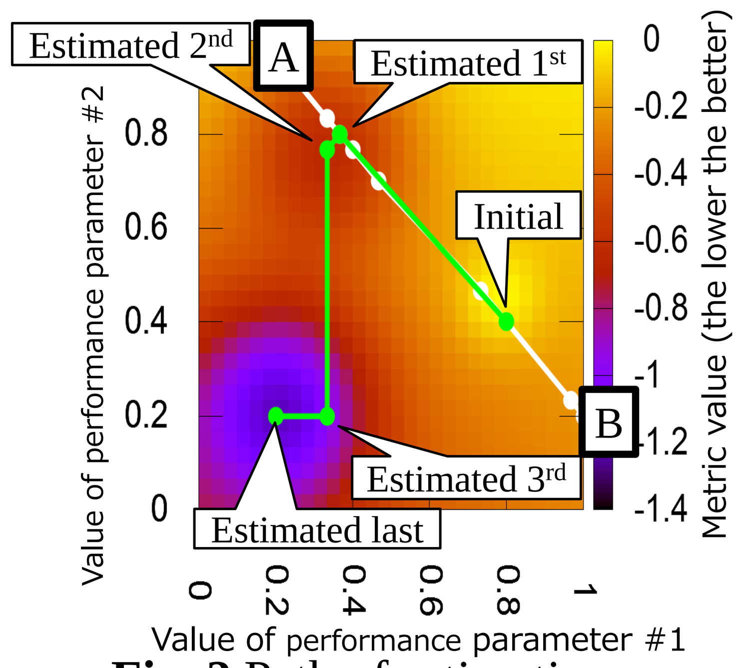

Fig. 2 Path of estimation

- Algorithm to estimate an optimum combination of performance parameters
  1. Directional exploration (Fig.3)
  2. Collinear exploration (Fig.4)
  3. Iterating 1. and 2. until convergence

✓ A darker point as a better combination
✓ A white line as a range of collinear exploration on Fig.4
✓ White points as the sampled combination on white line
✓ Green points as the estimated optimal combination
✓ Green lines as the path of estimated optimum points

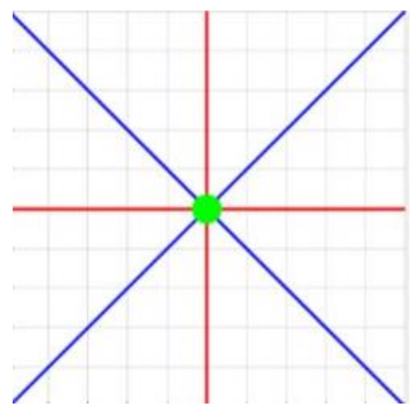

Fig.3 Determining a line[1]:
✓ A red line as a variable performance parameter
✓ A blue line as variable performance parameters

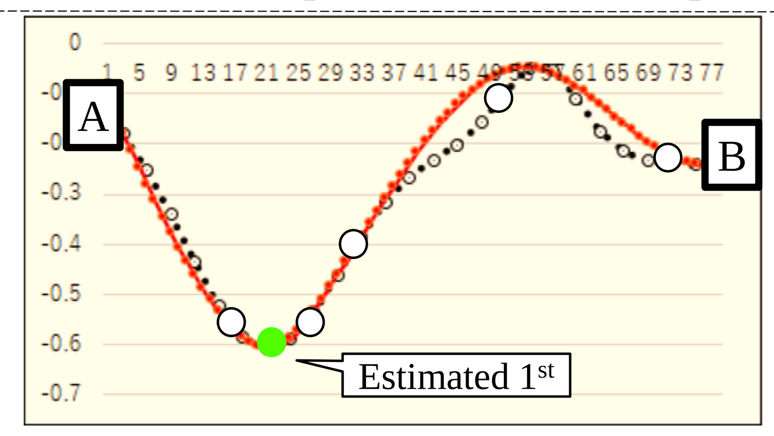

Fig.4 Fitting by d-Spline following some sampled metric value: Estimating the minimum metric value from A to B on Fig.2 [2].
✓ White points as each sampled value of combinations
✓ A red line as one-dimensional d-Spline following white points
✓ A green point as the value of the estimated 1st combination

The Application to Machine Learning

Estimating hyperparameters' configuration of a CNN model for an object classification problem called CIFAR10 reported in [4].

- Learning model : CNN
- Total patterns : 164,025

Table 1: Hyperparameters of CNN

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

◆ Metric value as Z considering both of classification accuracy and learning time
◆ $Z=(a \times L)+(b \times T)$
◆ L as accuracy loss
◆ T as normalized learning time
◆ a as weight for L
◆ b as weight for T
✓ a=1.0, b=0.5 in this experiment

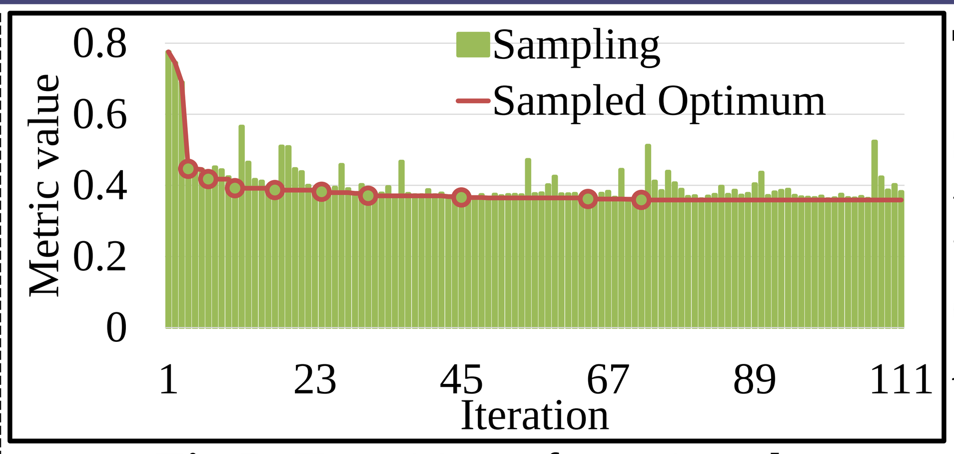

Fig.5: Transition of metric value estimated by DSICE

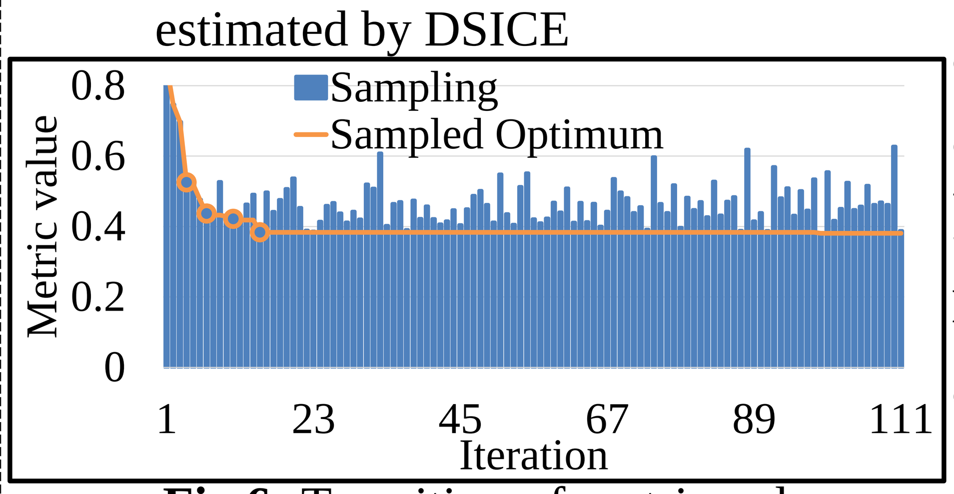

Fig.6: Transition of metric value estimated by Random Search

The estimated metric value are shown on Fig.5 by DSICE and on Fig.6 by Random Search. DSICE converges in 111 iterations for total 164,025 patterns. The accumulation of metric value is about 45 (green area) for DSICE, and about 50 (blue area) for Random Search. Thus DSICE reduces about 10% from the accumulation of metric value.

Conclusion

We develop a simple AT tool "DSICE" to estimate a combination of performance parameters for a target program, and apply DSICE to estimating hyperparameters' configuration of a CNN model. DSICE converges efficiently within approximately 100 iterations for 164,025 configuration patterns, and reduces 10% from the accumulation of metric value during estimation. We will apply DSICE to other performance optimization problems in our future work.

Acknowledgments

This study was partially supported by JSPS KAKENHI Grand Number JP 17K00164, JP 18K19782, JP 18K11340

Reference

[1] 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).
[2] 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.
[3] T. Katagiri,Auto-tuning forTheEra of Relatively High Bandwidth Memory Architectures:ADiscussion Based on an FDM Application, Proceedings of IEEE IPDPSW2018.
[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).