Parallelization of Automatic Tuning by Executing Machine Learning Programs in Multiple Jobs

Sorataro Fujika[†1*], Teruo Tanaka[†1], Akihiro Fujii[†1], Yuka Kato[†2], Satoshi Ohshima[†3], Takahiro Katagiri[†3]
†1 Kogakuin University, †2 Tokyo Woman's Christian University, †3 Nagoya University

Introduction

We are working on research to optimize hyperparameters by automatic tuning [1]. The machine learning program to be tuned takes several hours for one training, and it takes an enormous number of days for tuning. The purpose of this study is to parallelize the actual measurements required for automatic tuning and reduce the execution time.

Automatic tuning method

- We have proposed a method called iterative one-dimensional search[2].
- In the parameter space which consists of n kinds of parameters, the followings are repeated.
  1. The direction search to find the points around a certain point.
  2. The one-dimensional search to find the optimal point in the one-dimensional direction line.
- The number of search patterns in the entire parameter space is enormous.
  ⇨Start with narrow search direction and add the search direction sequentially.

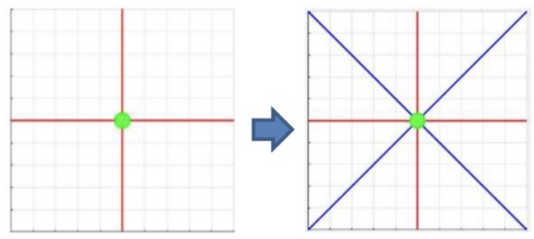

Fig.1 An Example of Expanding the Search Axis in a Two-Dimensional Search Space : Add a blue axis search to a red axis only search

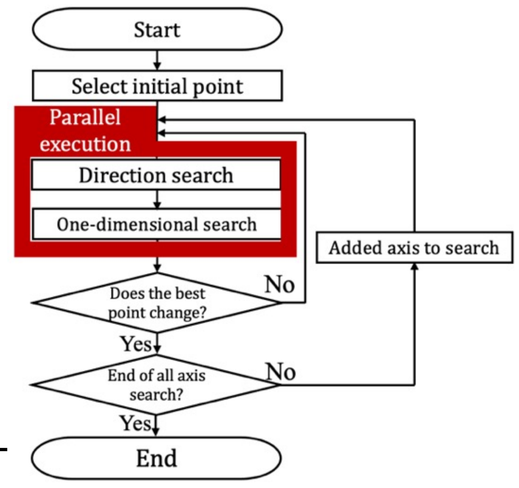

Fig.2 Perform directional search and one-dimensional search together

Parallelization method

- Hyperparameter estimation of a machine learning program takes a long time to do training, such as 35 days.
  ⇨The execution time is reduced by parallelizing the actual measurement of the direction search and the one-dimensional search.

The target program

- A pedestrian route prediction application using machine learning. [3]
  - ▸ Predicts the future route and arrival point of the pedestrian from the past movement trajectory data of the target person.
- Evaluation value : Final Displacement Error (FDE)
  - ▸ The error between the actual pedestrian's arrival point and the estimated arrival point.
  - ▸ The smaller the value, the better the accuracy.

Experiment outline

- Execution environment : “Flow” Type II subsystem
  - ▸ Supercomputer of Nagoya University
  - ▸ Multiple GPUs effective for machine learning can be used.
- Total pattern of parameters
  - ▸ $5^5 = 3125$ pattern

Table1. Parameter types and number of possible values

Hyperparameters	Number of parameters set
rnn_size	5
grad_clip	5
learning_rate	5
dropout	5
lambda_param	5

Estimated result

- Execution time is reduced to 1/14 by parallelization. (35days -> 2.4days)
- The estimated parameter has a 0.22m decrease in FDE compared to the initial parameter. (0.70m -> 0.48m)
  - ▸ The initial parameters the combination of hyperparameters at the start of search.
- Fig.3 show that 75% of the points searched this time have better FDE than green line.

- The blue dot : the FDE obtained as a result of executing the machine learning program.
- The area between the orange vertical lines : the range of simultaneous executions in parallel.
- The red dot : a point that becomes the smallest FDE found in that range.
- The notations for 1 to 5-axes : the number of axis orientations used for the search.
- The green line : the FDE value of the initial parameter.

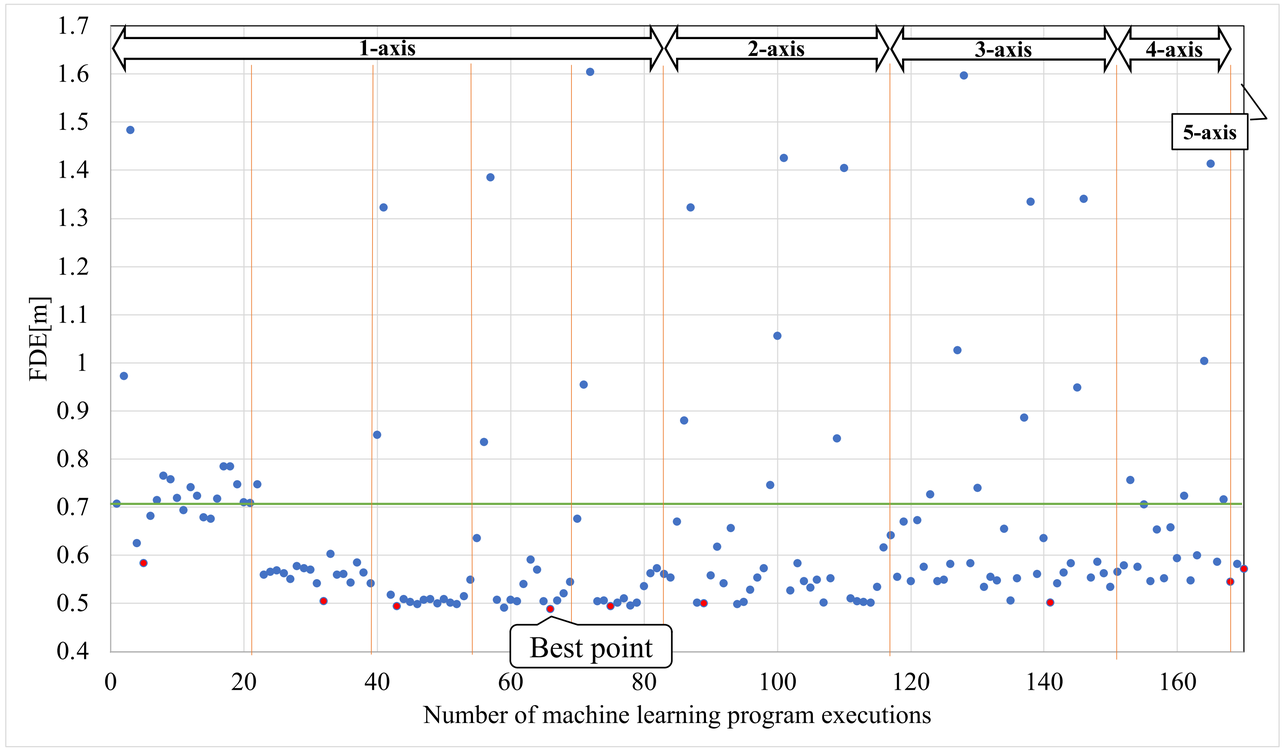

Fig3. Transition of search during parallel execution

Conclusion

In this research, an automatic tuning mechanism was made possible to measure in parallel by submitting multiple jobs at the same time. The conclusions of this report are that the time required for hyperparameter estimation of the target machine learning program was reduced from 35 days to 2.4 days by parallelization.

Acknowledgments

This work is supported by "Joint Usage/Research Center for Interdisciplinary Large-scale Information Infrastructures" in Japan (Project ID: jh210019-NAH) and JSPS KAKENHI Grand Number JP 18K11340.

Reference

[1] S. Fujika, T. Tabeta, A. Fujii, T. Tanaka, Y. Kato, S. Ohshima, T. Katagiri, Application and Stability Verification of Parallelized Automatic Tuning to Machine Learning Programs on GPU Cluster, IPSJ Technical Report, vol.2021-HPC-178, No.16, pp.1-8 (2021) (in Japanese).
[2] M. Mochizuki, A. Fujii, T. Tanaka, T. Katagiri, Fast Multidimensional Performance Parameter Estimation with Multiple One-dimensional d-Spline Parameter Search, International Workshop on Automatic Performance Tuning (iWAPT2017), (2017).
[3] R Akabane, Y Kato, Pedestrian Trajectory Prediction Based on Transfer Learning for Human-Following Mobile Robots, IEEE ACCESS, Vol.9, pp.126172-126185 (2021).