데이터베이스 튜닝을 위한 Knob 탐색 범위 축소 연구

권세인 ○ 이지은 박상현 †

연세대학교 컴퓨터과학과

seinkwon97@yonsei.ac.kr, jieun199624@yonsei.ac.kr, sanghyun@yonsei.ac.kr

A Study about Search Space of Knob Range Reduction for Database Tuning

Sein Kwon○ Jieun Lee Sanghyun Park†

Dept. of Computer Science, Yonsei University

ABSTRACT

Tuning methods to improve database performance are essential for efficiently processing large amounts of data. One methodology that can improve the performance of database is knob tuning methodology. However, since too many different types of knobs and a vast available range of knob values, the optimization algorithms are limited to explore the fine-grained knob values. To overcome this challenge, in this paper, we propose a methodology to reduce the search space of knob values within the optimization algorithm for database knobs tuning. This methodology can allow for more detailed exploration in optimization algorithm. Our various ablation studies show that knob tuning by using the narrow knob value performs better than knob tuning by default knob value range.

1. INTRODUCTION

Smart cities help citizens enjoy a convenient and comfortable life by solving various urban problems such as transportation, environment, and facilities. In particular, the Internet of Things in smart city services generates the vast amounts of data, so it is important to store, process and utilize big data [1].

However, managing large amounts of data requires a high-performance database. To improve the performance of an existing database efficiently, a database knob tuning can be considered. The database knob tuning is the process of adjusting knob settings to improve a database performance. However, database types are diverse and available knobs are different according to its version, it is not efficient that users manually tune the database knobs [2]. To address these issues, recent studies have focused on automatically recommending suitable knob values to improve database performance through optimization algorithms [3,4,5,6]. However, since databases contain numerous types of knobs, e.g., continuous, categorical and boolean, with a vast available range of values and the valid range changes depending on hardware specs, the optimization algorithms are limited to explore the fine-grained knob values, which is challenging to consider potential optimal values.

In this paper, we propose a methodology to reduce the search space of database knob values within an optimization algorithm. The proposed methodology defines a new range of knobs that can potentially enhance database performance, allowing more detailed exploration in the optimization algorithm and resulting in an enhancement of database performance.

To validate the proposed methodology, we conducted performance comparison experiments for throughput and latency with MySQL's default knob settings. Our results show that the range of knob values was significantly reduced, and the database performance was improved.

2. METHODOLOGIES

Unlike [3] that reduces a search space by selecting impactful knobs, our methodology is further aimed to reduce the value range of the impactful knobs. Figure 1 shows an overview of our proposed methodology. Firstly, we randomly generate 200 samples via Latin

† 교신 저자: sanghyun@yonsei.ac.kr

* 이 논문은 2023년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(IITP-2017-0-00477, (SW 스타랩) IoT 환경을 위한 고성능 플래시 메모리 스토리지 기반 인메모리 분산 DBMS 연구개발)과 국토교통부의 스마트시티 혁신인재육성사업으로 지원을 받아 수행된 연구임.

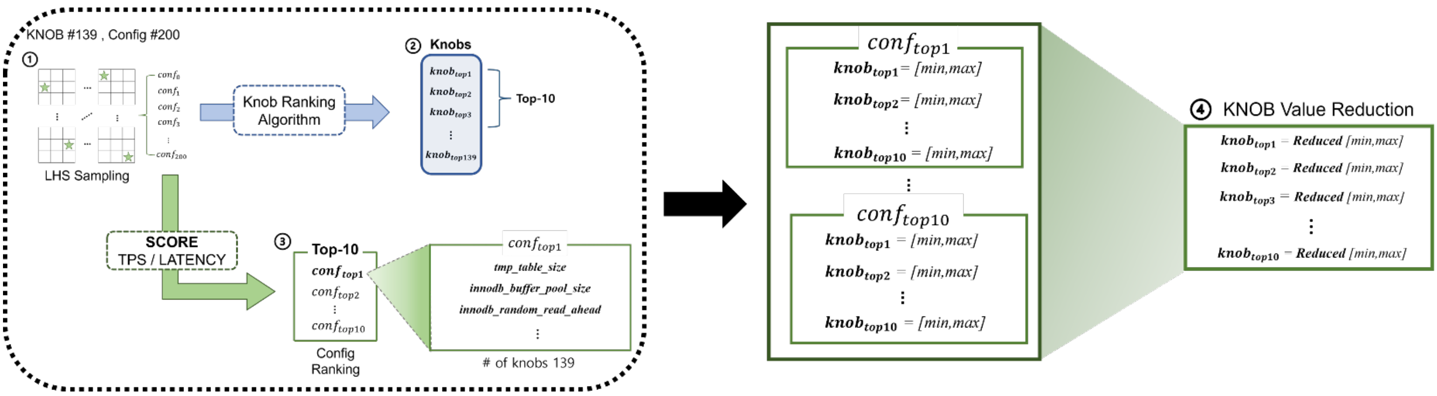

Figure 1. Design Overview of Proposed Methodology

Hypercube Sampling (LHS) [7]. Secondly, we select 10 knobs that have a significant impact on database performance by a knob ranking algorithm. Thirdly, 10 configurations within the generated samples are selected based on their measured database performance, where we calculated score (throughput/latency) to compare multiple configurations. Then, we find the used value range of each selected knob from the selected configurations. With these newly defined knob ranges, the optimization algorithm can search knob values within a narrower range than its default range.

3. EXPERIMENTAL EVALUATION

3.1 Experimental Setup

We conducted experimental evaluations in MySQL (v5.7) with 139 knobs using two workloads: TPC-C and Twitter.

To validate our methodology, we adopted Lasso and SHAP (Sharpley Additive exPlanations) as a knob ranking algorithm and BO (Bayesian Optimization) and SMAC as an optimization algorithm, which were applied to the existing studies [3,8,9,10]. We used XGBRegressor in SHAP and SMAC optimization model.

3.2 Evaluation

The experimental results of the ablation study are shown in Table 1 and Table 2. Table 1 is the results on the TPC-C workload using BO, and Table 2 is the results on the Twitter workload using SMAC. In each table, the results of our proposed methodology are denoted in bold. Figure 2 shows the default and reduced ranges of the knobs. In Figure 2, a thick bar represents the default range of knobs and a thin line above the bar indicates the reduced range by our proposed methodology. We also colored the bold bars differently depending on the amount of knob values to be explored.

We now demonstrate that our methodology can improve database performance. The results in Table 1 show that reducing both the number of knobs to be tuned and its range improves throughput by an average of 15.65% and latency by an average of 99.90% over the database performance with default MySQL knob settings. Furthermore, the results validate that our proposed methodology can improve the database performance even without limiting the number of knobs. The results in Table 2 also show that reducing the knob ranges improves tuning performance in the same way.

Table 1. Tuning Performance on TPC-C with Bayesian Optimization.

Knob Ranking	Range Reduction	Throughput (requests/second)	99th %-tile Latency (sec)
Default MySQL Setting		402.246	13.540
×	×	2.677	3.619
Lasso	×	0.033	56.011
SHAP	×	0.083	15.735
×	✓	405.883	0.019
Lasso	✓	**480.326**	**0.011**
SHAP	✓	**450.130**	**0.015**

Table 2. Tuning Performance on Twitter with SMAC.

Knob Ranking	Range Reduction	Throughput (requests/second)	99th %-tile Latency (ms)
Default MySQL Setting		6223.630	552
×	×	2009.716	594
Lasso	×	6987.064	561
SHAP	×	13.667	6514253
×	✓	6850.562	589
Lasso	✓	**7015.032**	**571**
SHAP	✓	**6894.030**	**570**

However, there are cases that achieved high performance without reducing the range of knob values

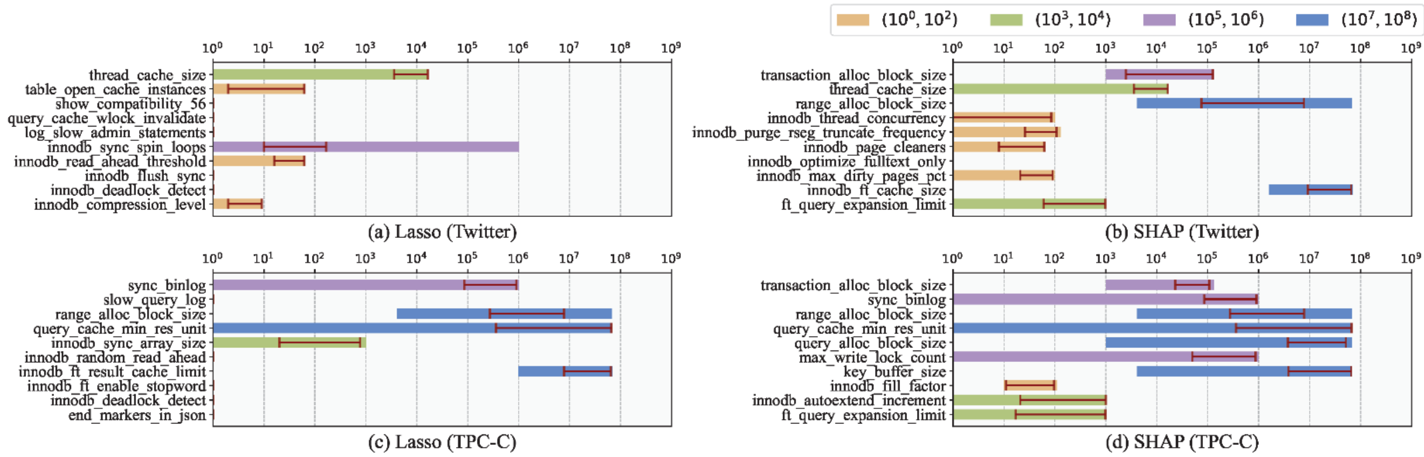

Figure 2. The default value range of the parameters extracted through the knob ranking algorithm and the value range after the proposed methodology.

in Table2. The reason is that the knobs in Figure 2 (a) consist of five knobs with boolean types and five knobs with continuous types. Compared to the knobs in Figure 2 (b), the search space for the knob values in Figure 2 (a) is much narrower. Specifically, we considered that the knob range with over 10^5 has a wide range. In Figure 2(a), there is only one knob with a wide range of 10^6, on the other hand, Figure 2(b) shows three knobs with wide ranges, where two knobs have much wider ranges of 10^7 to 10^8. This is why Figure 2(a) performs better than the default even without the knob reduction step.

In summary, we demonstrate that the process of defining the new range of knob values, which resulted in reducing the search space for knobs, is important for database knob tuning.

4. CONCLUSION

In database knob tuning research, there are some challenges such as a variety of knob types and the large range of continuous knob values. To reduce their large range, we proposed a methodology that selects the knobs that affect database performance through a knob ranking algorithm and then reduces the search space to the knob value range based on the configurations that have driven a high database performance. We verified our proposed methodology with experimental results compared to default knob settings.

In the future work, we will develop the proposed methodology to apply an online tuning for Cloud DBMS, which has drawn attention.

REFERENCES

[1] Glasmeier, Amy, and Susan Christopherson. "Thinking about smart cities." *Cambridge Journal of Regions, Economy and Society* 8.1 (2015): 3-12.

[2] 권세인, JIN HUIJUN, 박상현. (2022).데이터베이스 파라미터 튜닝 기법 비교 분석.한국정보과학회 학술발표논문집,(),98-100.

[3] Van Aken, Dana, et al. "Automatic database management system tuning through large-scale machine learning." *Proceedings of the 2017 ACM international* conference on management of data. 2017.

[4] Zhang, Ji, et al. "An end-to-end automatic cloud database tuning system using deep reinforcement learning." *Proceedings of the 2019 International Conference on Management of Data*. 2019.

[5] Khan, Md Muhib, and Weikuan Yu. "Robotune: high-dimensional configuration tuning for cluster-based data analytics." *50th International Conference on Parallel Processing*. 2021.

[6] Fekry, Ayat, et al. "Tuneful: An online significance-aware configuration tuner for big data analytics." *arXiv preprint arXiv:2001.08002* (2020).

[7] Iman, Ronald L. "Latin hypercube sampling." Encyclopedia of quantitative risk analysis and assessment 3 (2008).

[8] Duan, Songyun, Vamsidhar Thummala, and Shivnath Babu. "Tuning database configuration parameters with ituned." Proceedings of the VLDB Endowment 2.1 (2009): 1246-1257.

[9] Fekry, Ayat, et al. "To tune or not to tune? in search of optimal configurations for data analytics." Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2020.

[10] Zhang, Xinyi, et al. "Restune: Resource oriented tuning boosted by meta-learning for cloud databases." Proceedings of the 2021 international conference on management of data. 2021.