

DPPML : 메타러닝을 활용한 데이터베이스 성능 예측

DPPML : Database Performance Prediction with Meta-Learning

연세대학교 컴퓨터과학과 염찬호

2022년 07월



과제명: IoT 환경을 위한 고성능 플래시 메모리
스토리지 기반 인메모리 분산 DBMS 연구개발

과제번호: 2017-0-00477



과학기술정보통신부
Ministry of Science and ICT



연세대학교
YONSEI UNIVERSITY



정보통신기술진흥센터
Institute for Information & communications Technology Promotion

Content

DPPML : 메타러닝을 활용한 데이터
베이스 성능 예측

- Introduction
- Method
- Experiment & result
- Conclusion

Introduction

- 데이터 베이스의 Knob 설정에 따라 데이터베이스 성능 상이
→ 데이터베이스 Knob 튜닝 필요
- 데이터베이스 Knob 튜닝 시 빠르고 신뢰성 있게 성능 예측하는 모델 필요.
+ 같은 Knob 설정이라도 어떤 워크로드에서 행해지는지에 따라 성능 상이
→ 워크로드 환경에 따라 달라지는 성능을 예측할 수 있는 모델 필요

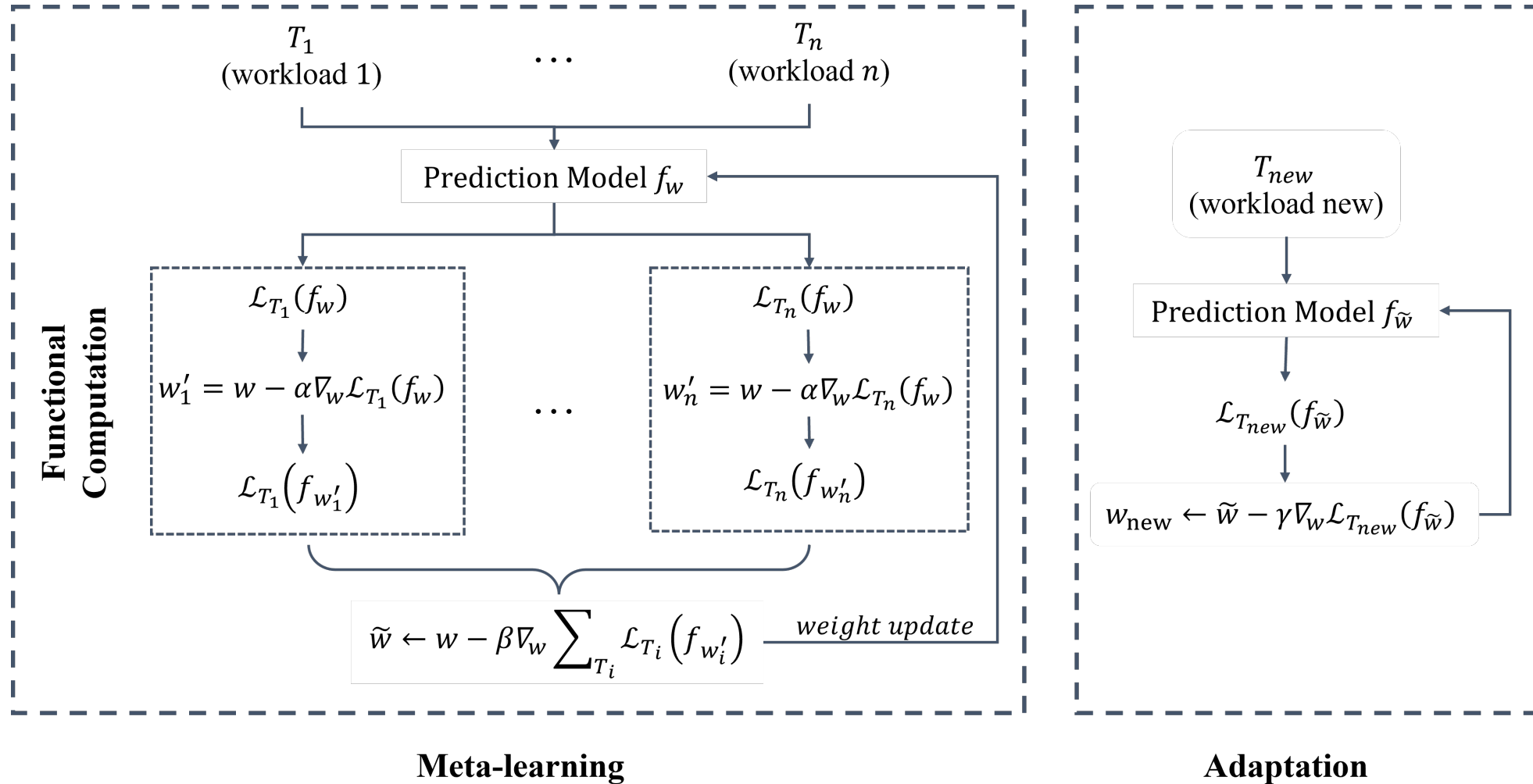
Introduction

- 일반적인 모델은 새로운 워크로드 환경이 주어졌을 때 예측 정확도가 낮음
- 새로운 워크로드 환경에 대해 학습하기 위해서 많은 양의 데이터가 추가로 필요

→ 새로운 워크로드 환경에 대해서 적은 데이터로 빠르게 수렴하고 예측 정확도가 높은 DPPML 모델을 제안

Method

- 전체 모델 구조



Method

- 일반적인 학습과정

Dataset : D

Model : f



$\mathcal{L}_D(f_w)$



$$w \leftarrow w - \alpha \nabla_w \mathcal{L}_D(f_w)$$



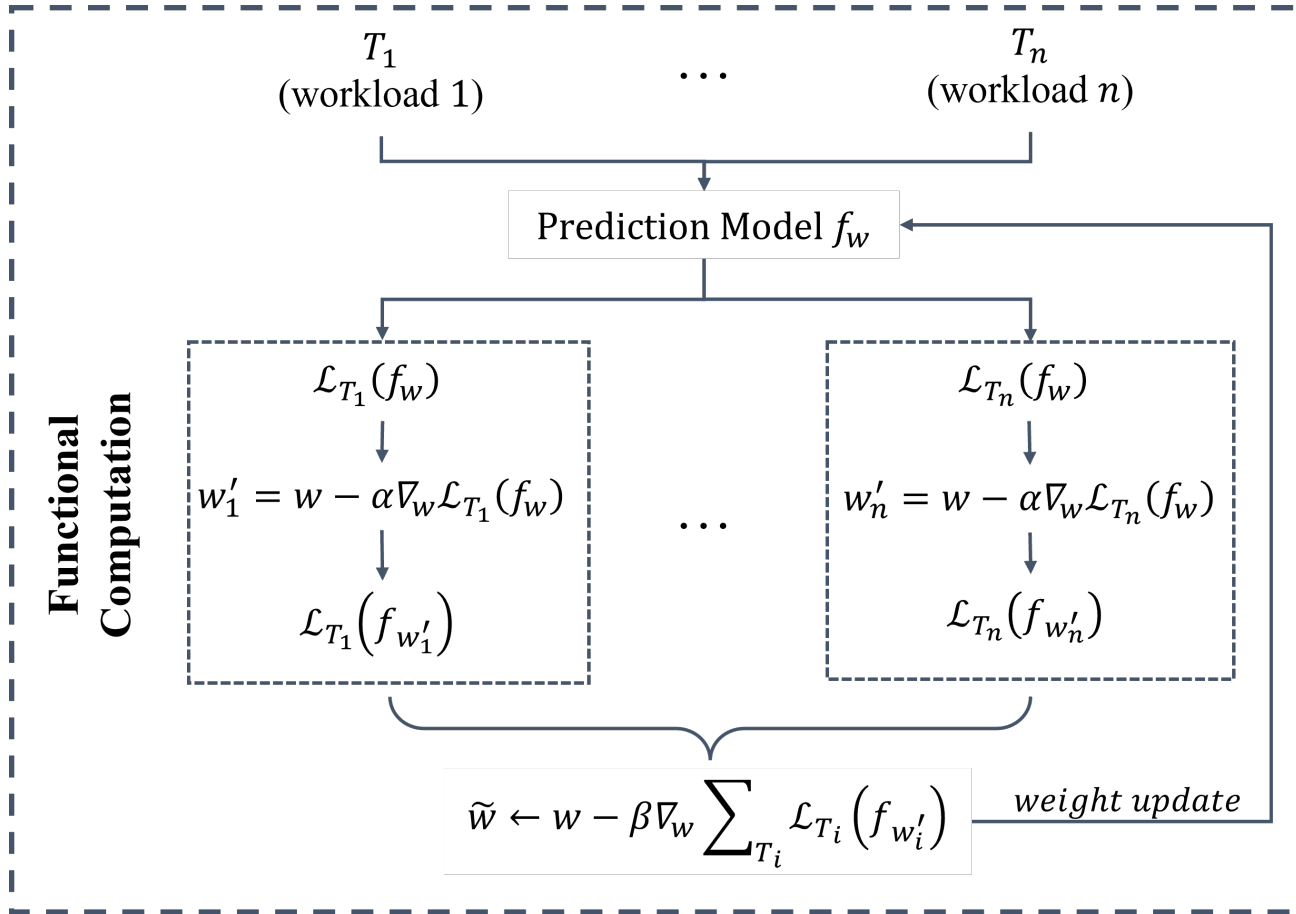
$\mathcal{L}()$: 손실함수(loss function)

w : f 의 가중치(weight)

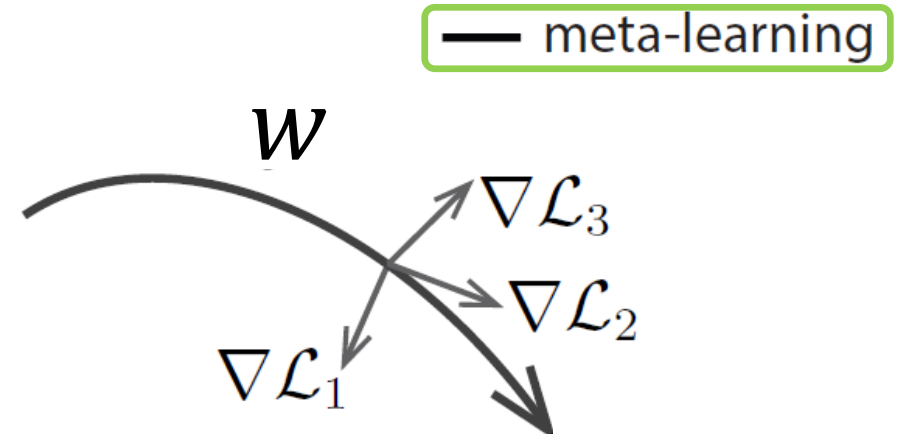
α : 학습률(learning rate)

Method

• Meta-learning

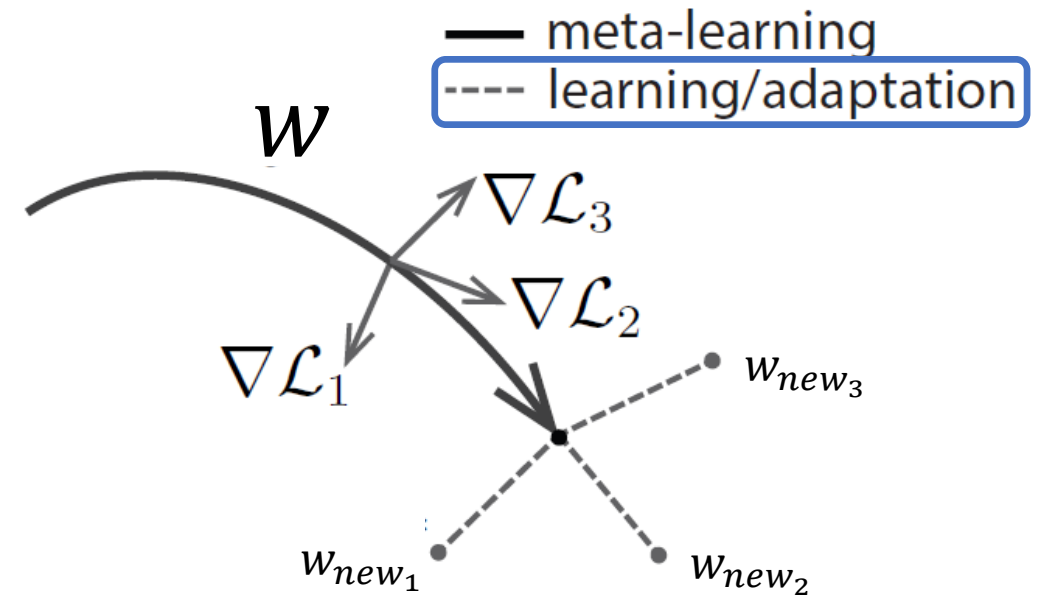
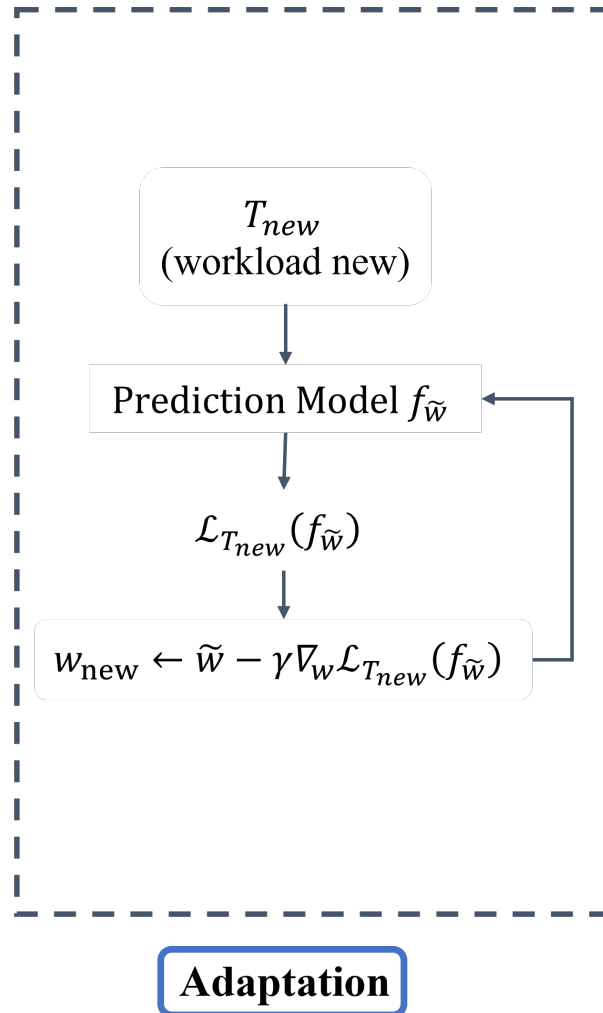


Meta-learning



Method

- Adaptation



Experiment

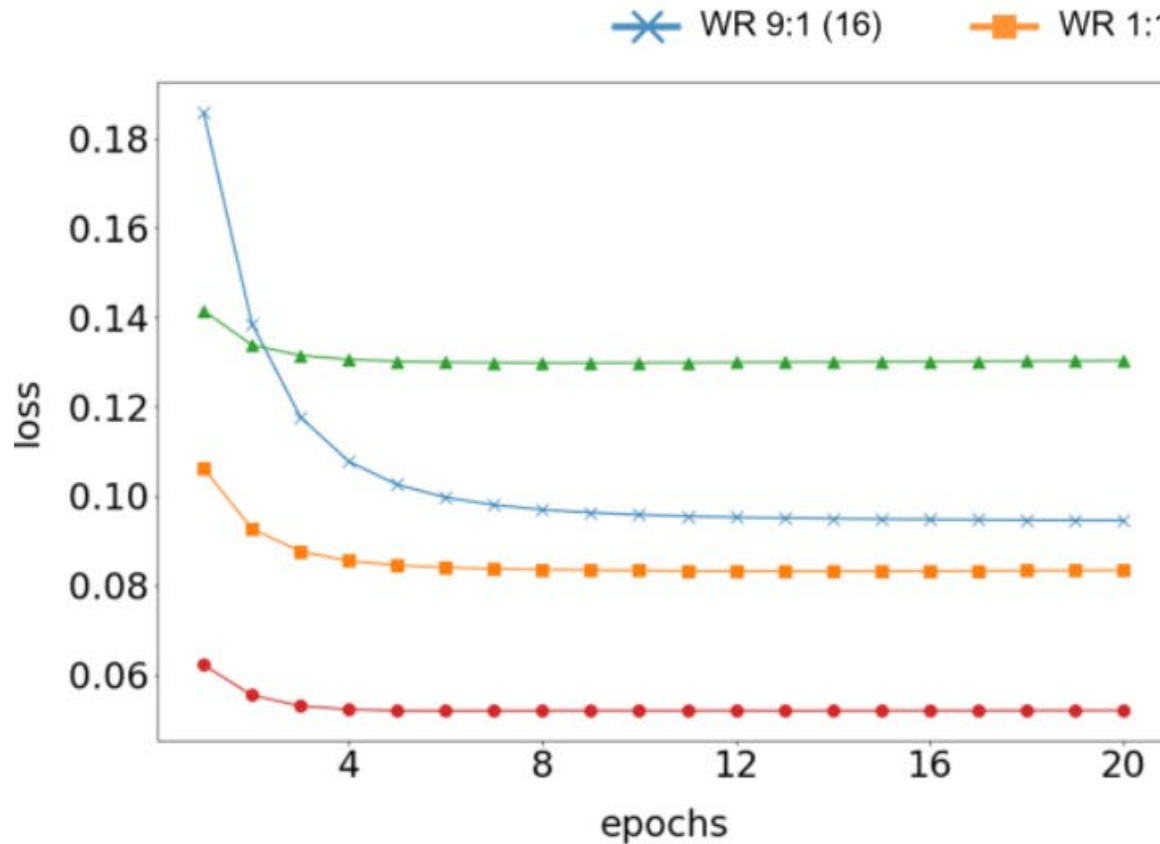
- 실험환경
- 워크로드 환경

Meta-learning dataset			
RW 9:1 (1)	RW 1:1 (1)	RW 1:9 (1)	Update (1)
RW 9:1 (4)	RW 1:1 (4)	RW 1:9 (4)	Update (4)
RW 9:1 (64)	RW 1:1 (64)	RW 1:9 (64)	Update (64)
Adaptation dataset			
RW 9:1 (16)	RW 1:1 (16)	RW 1:9 (16)	Update (16)

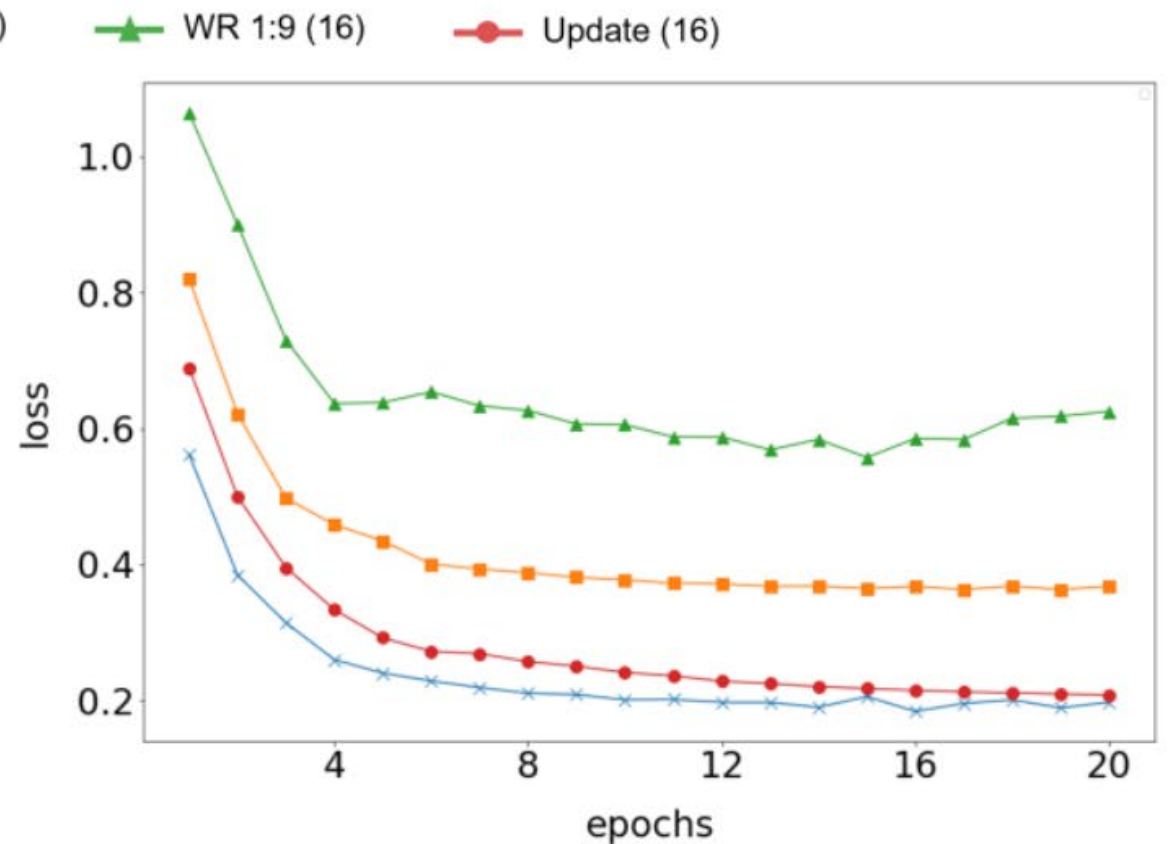
- 실험 데이터셋

RocksDB의 Knob과 워크로드에 대해 db_bench로 성능 측정한 데이터

Result



(1) DPPML



(2) Pretrained

Adaptation과정에서 DPPML과 Pretrained모델의 epoch당 loss(손실함수 값) 그래프

Result

Model	# of Samples	RW 9:1 (16)	RW 1:1 (16)	RW 1:9 (16)	Update (16)
DPPML	10	0.6646	0.8598	0.8406	0.8995
	40	0.6970	0.8676	0.8734	0.9148
	70	0.7103	0.8709	0.8413	0.9120
	100	0.7025	0.8703	0.8759	0.9151
Pretrained	10	0.4698	0.4363	0.5746	0.6681
	40	0.4478	0.5877	0.5967	0.6046
	70	0.4711	0.6587	0.6674	0.6886
	100	0.4981	0.6980	0.7382	0.7560

데이터 샘플 개수에 따른 R2 score(↑) 비교

Conclusion

- 데이터베이스 Knob 튜닝 시 같은 Knob 설정에 대해서도 어떤 워크로드에서 수행되는지에 따라 결과가 상이함.
- 일반적인 모델은 새로운 워크로드 환경이 주어졌을 때 예측 정확도가 낮음
- 새로운 워크로드 환경에 대해 학습하기 위해서 많은 양의 데이터가 추가로 필요
- 본 논문에서 제안하는 메타러닝을 접목시킨 DPPML은 적은 데이터로 빠른 속도로 수렴하고 예측 정확도가 높음

Thank You !