Check-N-Run: a Checkpointing System for Training Deep Learning Recommendation Models

Assaf Eisenman, Kiran Kumar Matam, Steven Ingram, Dheevatsa Mudigere, Raghuraman Krishnamoorthi, Krishnakumar Nair, and Misha Smelyanskiy, Facebook; Murali Annavaram, Facebook and USC

USENIX NSDI'22

September 13th, 2022

Presented by Yejin Han

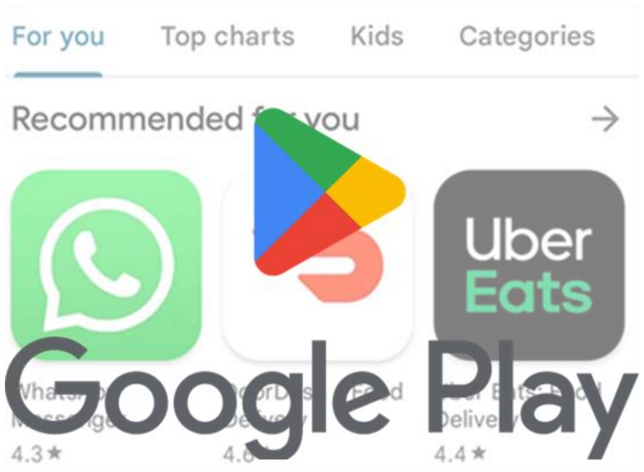
yj0225@dankook.ac.kr

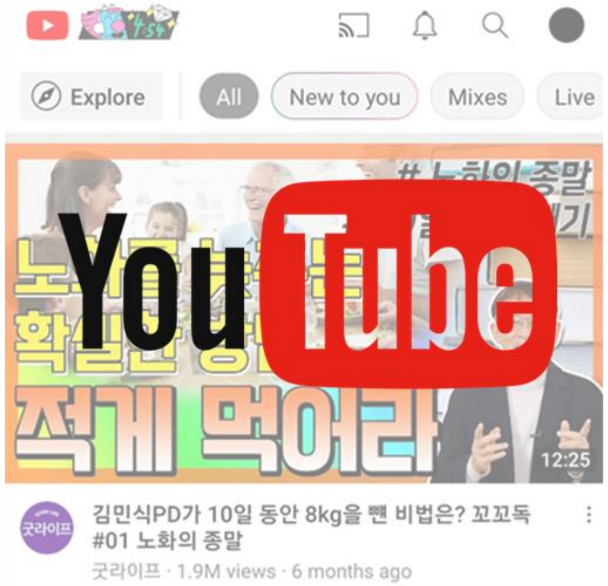
Introduction

- Personalized recommendation is everywhere

Introduction

- Recommendation models are important
  - 40% of apps installs on Google Play
  - 60% of watch time on YouTube
  - 35% of purchase on Amazon
  - 75% of movie watches on Netflix





What is Recommendation Model?

- Recommendation model
  - Updates only a small fraction of the model during each iteration
  - Exceed Terabytes, which stress planetary scale storage system

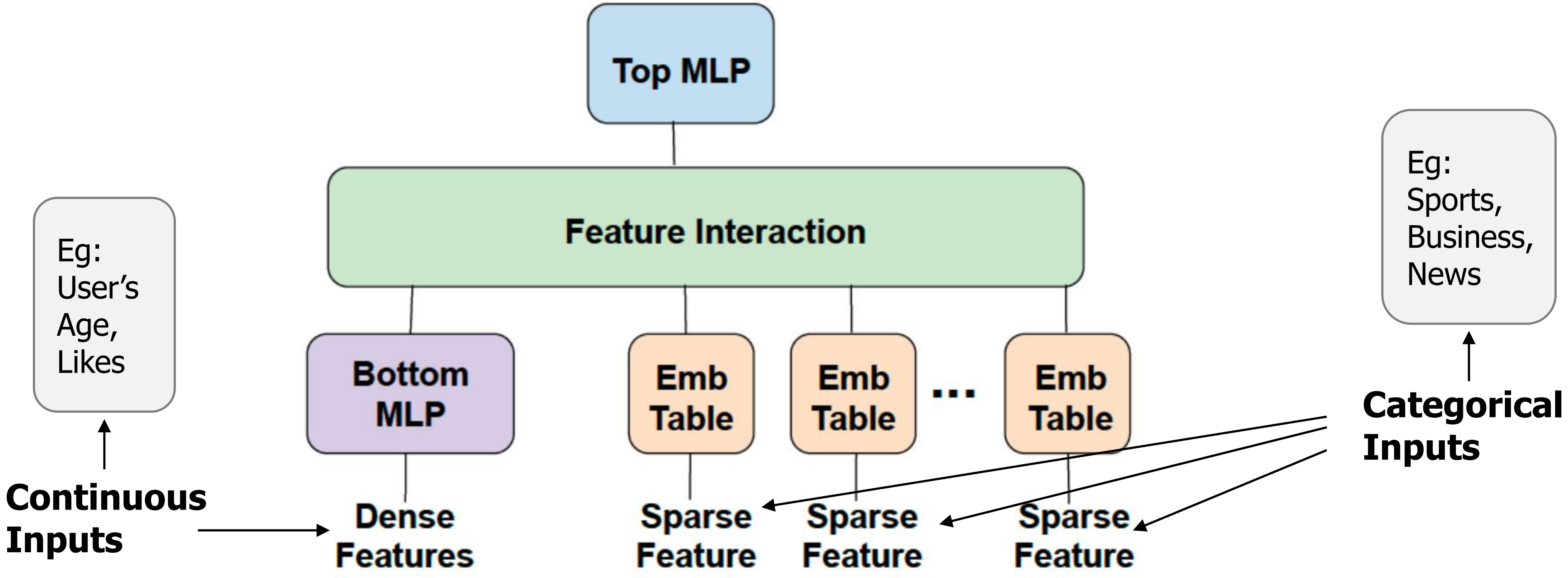


Figure 1: A typical recommendation model

Background

- High performance training at Facebook

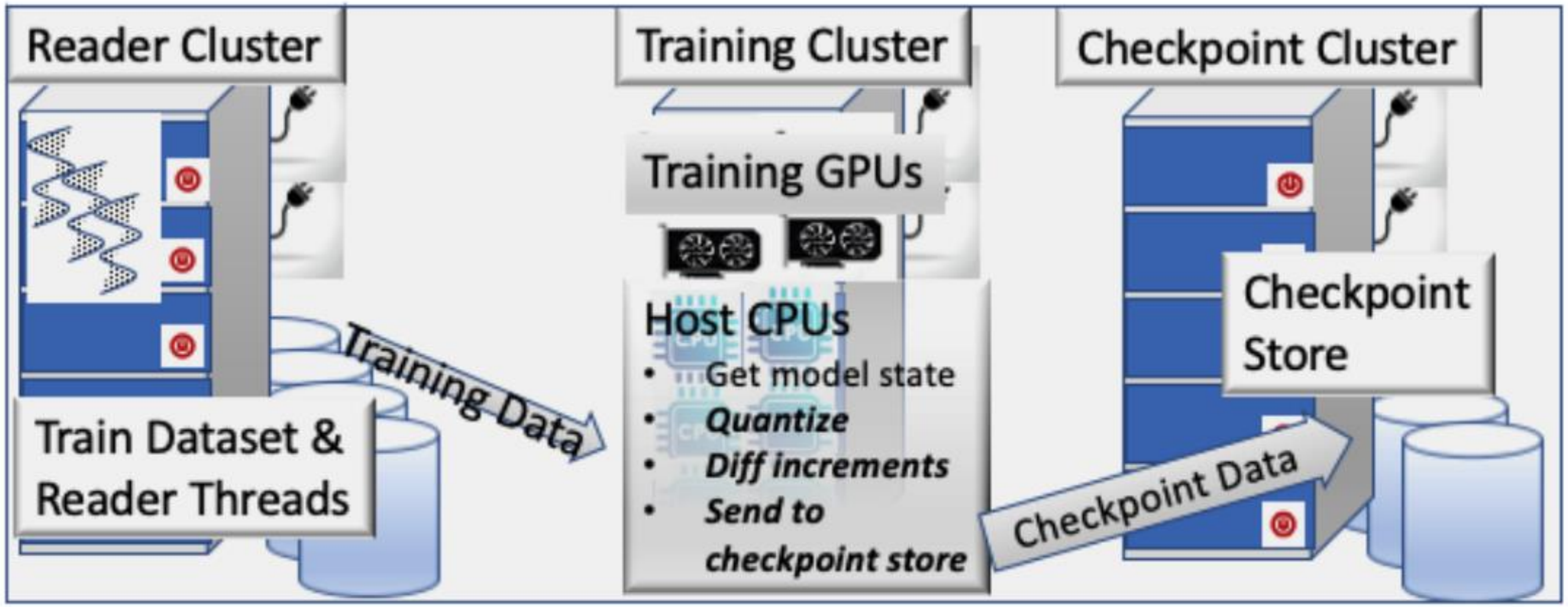


Figure 2: An Overview of Training and Checkpoint Systems

Why do we need checkpoints?

- Failure recovery
- Migrating training processes
- Publishing snapshots in real time
- Transfer learning

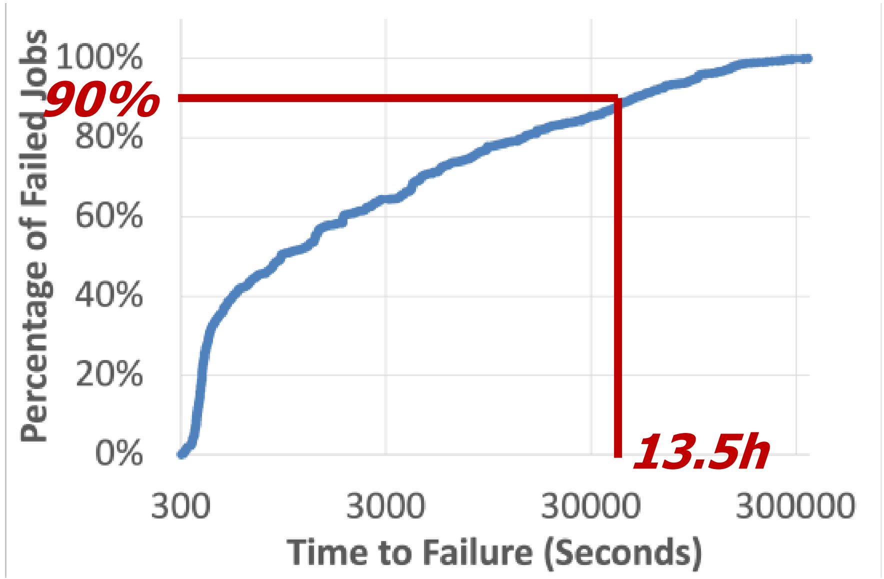


Figure 3: Training job failure CDF in our cluster. Jobs that fail within 5 minutes are removed since they are usually simple user setup errors.

What are the checkpoint challenges?

- Accuracy
- Frequency
- Write bandwidth
- Storage capacity

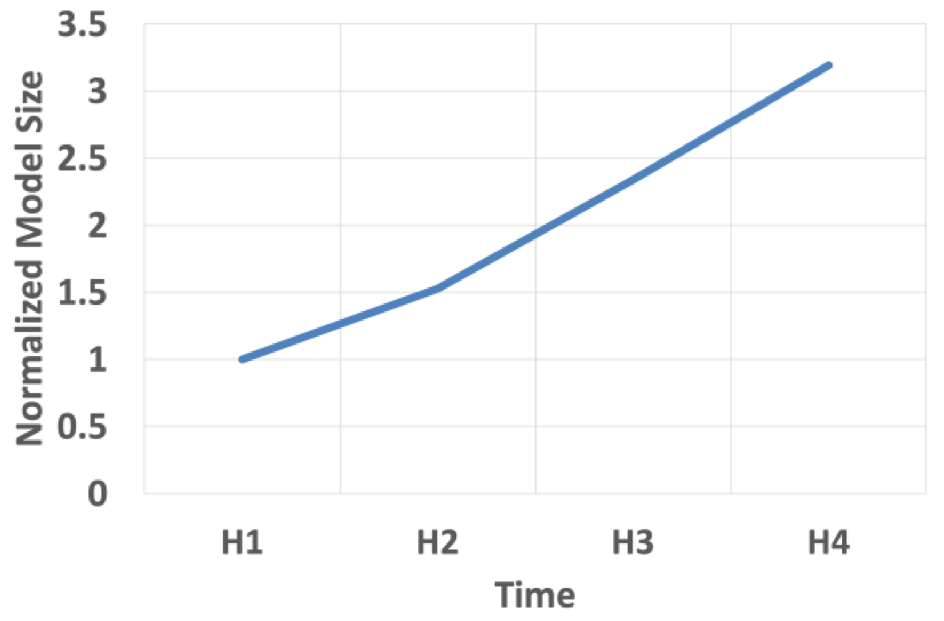


Figure 4: The normalized model size over the past 2 years

Check-N-Run

- Checkpointing workflow

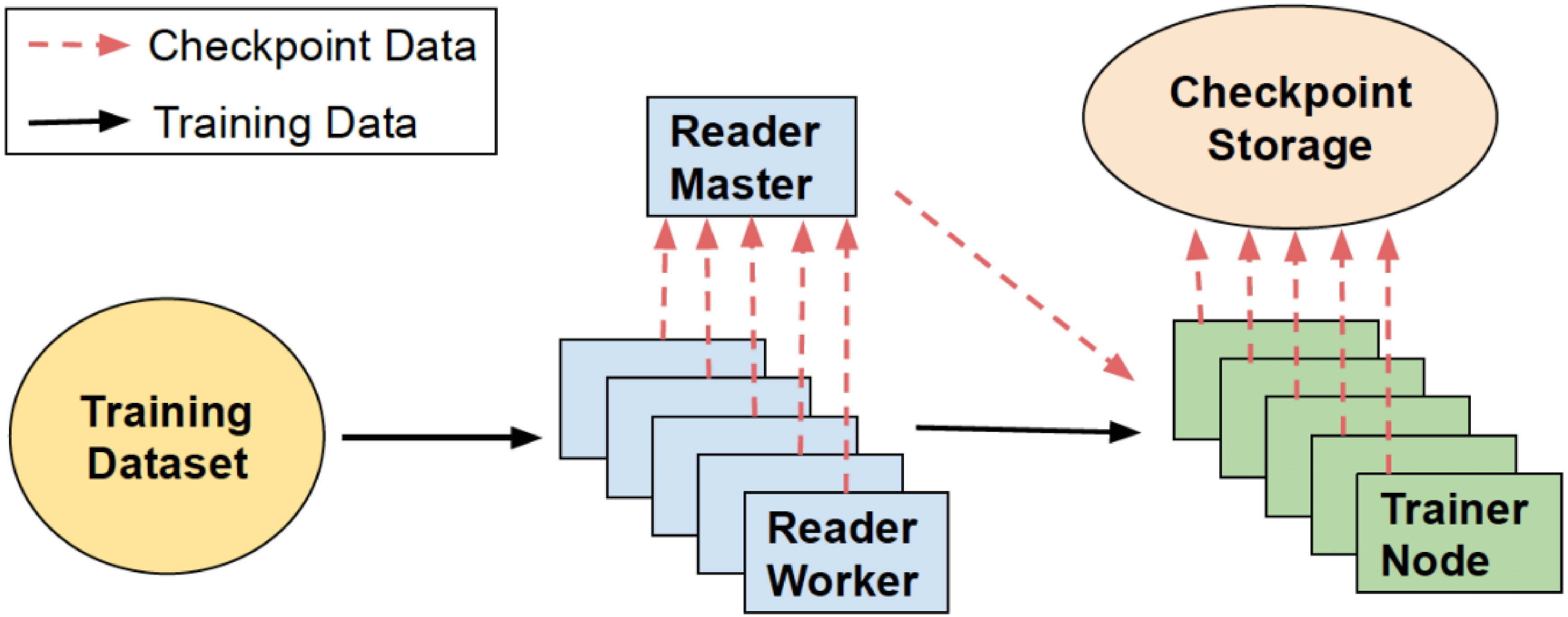


Figure 8: High-level data flow during training

Differential Checkpointing

- One-Shot Differential Checkpointing
  - Takes single checkpointing as a full baseline checkpoint
  - Stores modified vectors since the baseline checkpoint

- Consecutive Incremental Checkpointing
  - Stores the vectors that were modified only during the last interval
  - Requires keeping all previous incremental checkpoints for reconstructing the model

- Intermittent Differential Checkpointing
  - Does a full model checkpointing intermittently
  - Decides when to take a full checkpoint using a simple history based predictor

Evaluation

- Differential Checkpointing Policy Comparison

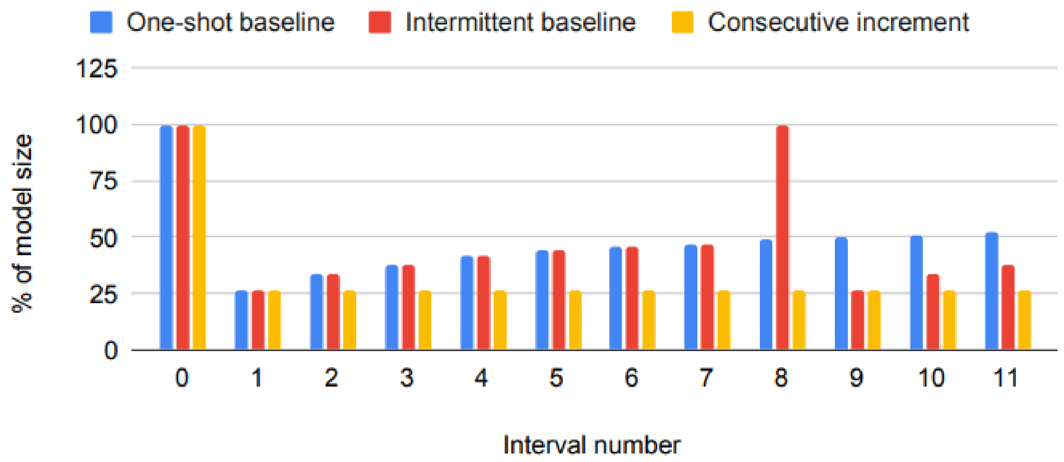


Figure 15: Bandwidth measure: checkpoint size per interval of 30 minutes, for different checkpoint policies

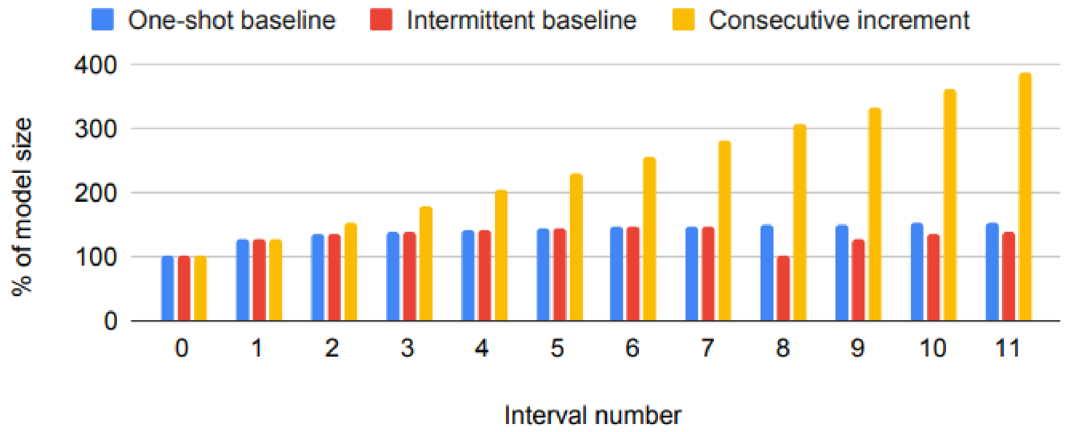
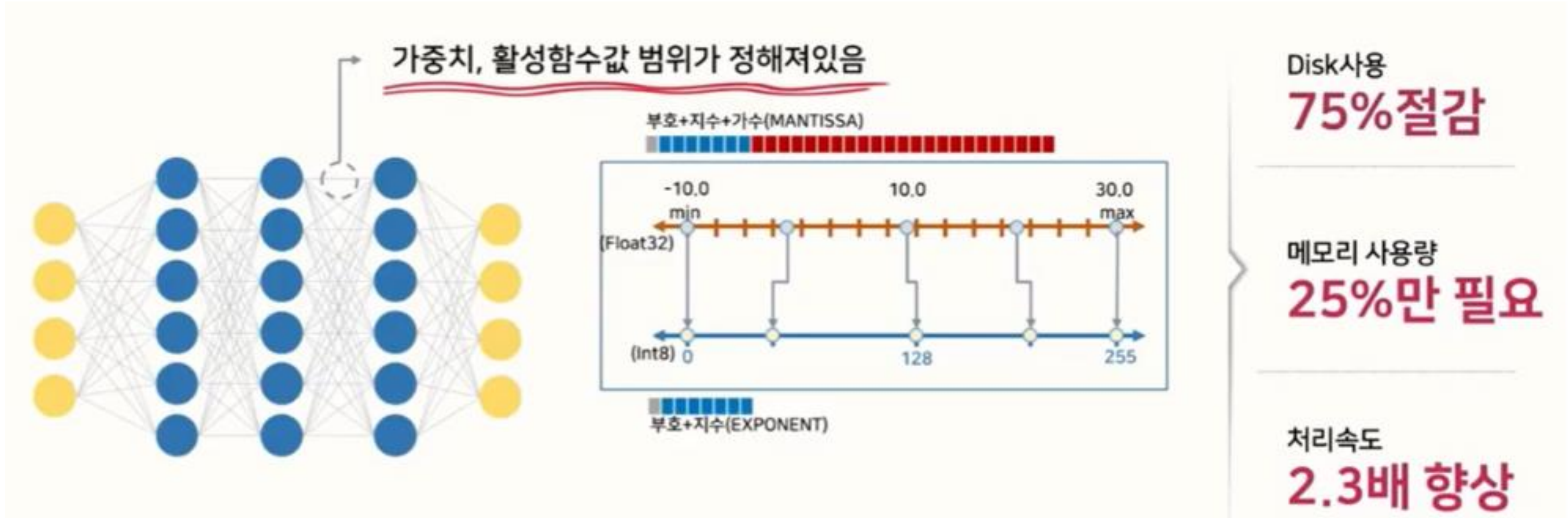


Figure 16: Storage measure: the required storage capacity at each interval of 30 minutes, for different checkpoint policies

Checkpoint quantization

- Compress checkpoint without degrading training accuracy



(source: https://gaussian37.github.io/dl-concept-quantization/)

Quantization strategies

- Uniform quantization
  - Symmetric, Asymmetric uniform quantization
- Non-uniform quantization using k-means
  - Embedding vectors are not all mapped into equally spaced buckets
- Adaptive uniform quantization
  - Leverage a greedy search algorithm to to select the Xmin and Xmax values

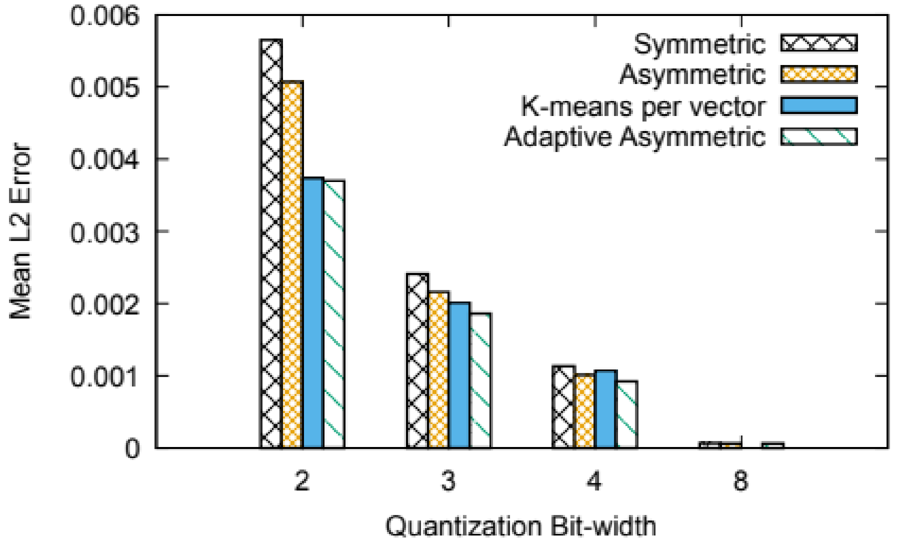


Figure 9: Mean ℓ_2 error of a quantized checkpoint for different quantization approaches

Evaluation

- The training accuracy

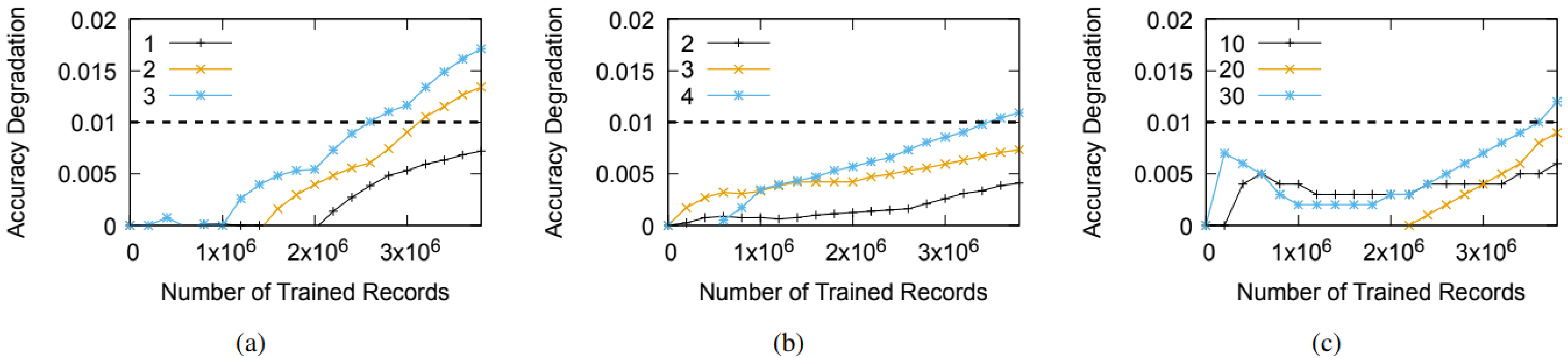


Figure 14: Lifetime accuracy degradation in a training job of 4 billion training samples, when using: (a) 2-bit, (b) 3-bit, and (c) 4bit quantized checkpoints. The lines represent the number of times the job had to resume from a quantized checkpoint

Evaluation

- Overall reduction in write bandwidth and storage capacity

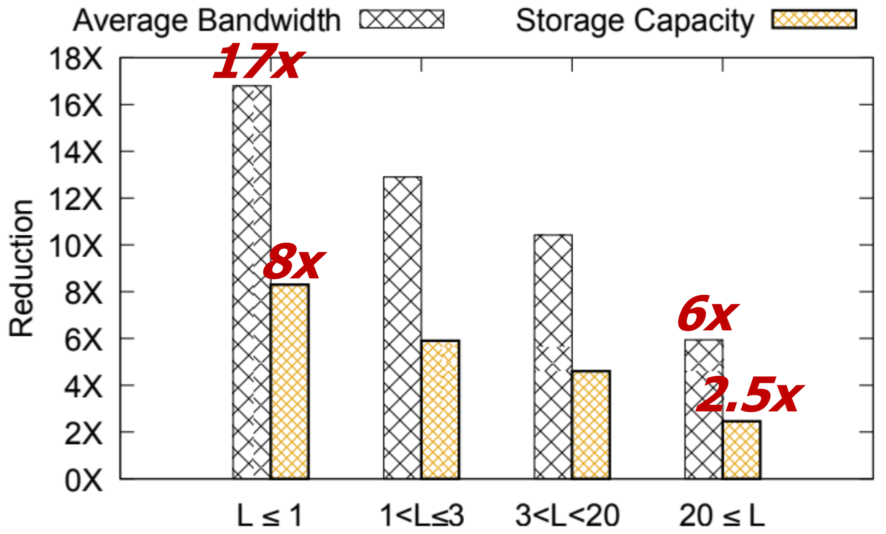


Figure 17: Overall reduction of the checkpoint average write bandwidth and storage capacity. *L* represents the number of times the training job had to resume from a checkpoint.

Conclusion

- The checkpointing of large recommendation systems at scale is challenging
- Check-N-Run uses strategies of differential checkpointing and quantization
- Check-N-Run significantly reduces the required write-bandwidth and storage capacity

Check-N-Run: a Checkpointing System for Training Deep Learning Recommendation Models

Assaf Eisenman, Kiran Kumar Matam, Steven Ingram, Dheevatsa Mudigere, Raghuraman Krishnamoorthi, Krishnakumar Nair, and Misha Smelyanskiy, Facebook; Murali Annavaram, Facebook and USC

USENIX NSDI'22

Thank You !

September 16th, 2022

Presented by Yejin Han

yj0225@dankook.ac.kr