Learning Algorithms for Automatic Data Structure Design

Demi Guo | Advised by Subarna Chatterjee and Stratos Idreos

Storage Engines are Used in Numerous and Diverse Applications

Different Storage Engines have Different Performance Properties

Design Continuum Introduces A Way to Decompose SE Design Space

How to Search in the Vast and Continuous Design Space?

θ: parameters of neural cost model
W: workload
H: hardware spec
D: design knobs

\[ \min_{\theta} \mathbb{E}_{W,H,D}[f_{\text{neural}}(W,H,D; \theta) - f_{\text{real or analytical}}(W,H,D)]^2 \]

(1) Train a Neural Cost Model on Constructed Dataset - a Differentiable Cost Model Easier to Optimize
(2) Gradient-Descent Based Traversal in the Search Space - Find Best Designs which Minimizes Neural Cost Model

Better or equivalent solutions with better search efficiency
(30% better designs, up to 20% lower I/O cost on average, 14x faster)

Discussion & Future Work

Results & Discussion