

SQUIDにおける分散深層学習のスケーラビリティ向上

大阪大学 D3センター 高橋慧智

- 目的** SQUIDにおける分散深層学習処理のスケーラビリティを測定・改善する。
- 内容** MLPerf HPCのCosmoFlowベンチマークをSQUID GPUノード上で実行し，分散深層学習処理を大規模実行した際のボトルネックを特定・緩和した。
- 結果** 大規模実行時にはストレージI/Oおよび集団通信がボトルネックになることを特定し，データセット読み込みの並列化，データセット前処理のGPUオフローディング，データセットのGPUメモリへのプリフェッチ，CPUコア割当の固定などの最適化を実施した。その結果，1ノードで3.6倍，16ノードで2.4倍の性能向上を達成できた。

	SQUID GPUノード
ノード時間	512ノード時間
使用メモリ	~9TiB
並列化	~32ノード (256GPU) Horovod+MPI+NCCLを使用

