

【第6回】 富士通次世代 コンピューティングセミナー

大規模言語モデル(LLM)の学習を支える分散学習技術

本田 巧/Honda Takumi

(富士通株式会社コンピューティング研究所 シニアリサーチマネージャー)

- Transformerアーキテクチャや大規模言語モデル開発の背景にあるScaling Lawの紹介
- GPT-3を例にLLMの学習に必要な3次元の並列化手法や高速化技術について解説

近似コンピューティングで計算機システムの高速化と省電力化を両立しよう！！

鯉渕 道紘/Koibuchi Michihiro

(国立情報学研究所／総合研究大学院大学 教授)

- アプリケーションが要求する「正確さ」とシステムが供給する「正確さ」の溝を埋める
- プロセッサ・メモリ・ネットワークを例に計算機システムの性能を改善する可能性を描く

2024年03月04日 (月) 14:00-16:00

【ハイブリッド開催】

オフライン：広島大学情報メディア教育研究センター本館2階 (申込不要)

オンライン：Microsoft Teams会議 (QRコードから詳細確認・参加登録) →

【主催】 富士通次世代コンピューティング共同研究講座

【お問い合わせ】 tsuji@hiroshima-u.ac.jp

