

階層ディリクレ過程のハイブリッド並列化推定

大元 司 (神戸大), 江口 浩二 (神戸大), 東羅 翔太郎 (NTT)

目的

近年ますます大規模化する情報(データ)を効率的に分析する手法を提案する

- ノンパラベイズアプローチをLDAに適用したHDP-LDAの並列化推定
- 通信コストの削減
- メモリ使用量の削減

ハイブリッド推定手法

- AD-HDP [Newman+,2009]などをベースに
- ノード間: MPI
- ノード内: 共有メモリ

実験

- 比較的小さなデータ (文書数: 1,500)
- より大規模なデータ (文書数: 39,861)

