

Research on Accelerator for Bayesian Inference

Name : Hiroki Nishimoto
Laboratory : Computing Architecture Lab.
Supervisor : Prof. Yasuhiko Nakashima

Abstract

ビッグデータ分析の手法として、ニューラルネットワークを用いる他に、データを数理モデルとして定義し、データを生成する確率分布を設定し、ベイズ推定を基礎とするアルゴリズムにより、データから確率分布のパラメータを推論する手法がある。ベイズ推定を用いた手法はニューラルネットワークを用いるものより説明性が高く、ニューラルネットワークが広く用いられるようになった現在でも、データ解析の主要技術として使われている。本研究は、二部構成である。

第一部は、代表的なビッグデータ分析手法の1つとして、様々な性質や特徴が混ざりあう中から、類似性によりグループに分類し属性を分析する、クラスタリングと呼ぶ手法の高速化に関する。クラスタリングは、ユーザーのセグメント分析やブランドポジションの認知等の手段として着目される手法である。実現手法は様々あり、中でも、混合ガウス分布を用いる手法は精度が高い。さらに、ガウス分布に共役事前分布を導入した変分推論法は、過学習が発生し難い特長があり、また、クラスタ数を自動決定できる。しかし、混合ガウス分布に関する変分推論法は、収束までに時間がかかることが知られおり、データ数に比例して、計算時間は膨大となる。そこで、変分混合ガウス分布を大規模データに適用するために、並列処理系である GPU を用いて、変分推論法の実装および評価を行った。提案手法は、同アルゴリズムの CPU 実装と比較して、同精度を保ちつつ、データ数が約 100 万個のとき 192 倍、クラスタ数が 256 個のとき 107 倍に、それぞれ高速化できた。また、EM アルゴリズムを用いる手法と比較して、同等のスピードを維持しつつ、クラスタ数の過度な縮退を防ぐことができた。

第二部は、逐次モンテカルロ法において、パラメータ推定を高速化するハードウェアの設計・評価に関する。前述の変分推論法は、強力なパラメータ推定手法である。しかし、モデルや使用する確率分布によっては適用範囲が狭く、変分混合ガウス分布等、計算アルゴリズムが確立されているモデルも少ない。さらに、複雑なモデルは実装困難である。代わりに用いられるのが、マルコフ連鎖モンテカルロ法に代表される、乱数生成器を用いるパラメータ推定手法である。乱数生成を用いる手法は、データの事前分布から乱数を生成し、データとの尤度を計算し、尤度によってパラメータを最適化する手法である。逐次モンテカルロ法はパーティクルと呼ぶ、多くの確率分布群を生成する。そして、各群から乱数を生成し、データに対する尤度を計算し、尤度の高いものをコピーし、低いものを棄却するという、ハードウェアレベルの並列化に適したものである。本研究では、逐次モンテカルロ法によるパラメータ推定高速化のため、ボトルネックである、リサンプリングステップと呼ぶ過程のハードウェア化に関するものである。メトロポリス・リサンプリング専用ハードウェアの効率向上を目的とし、固定小数点等、様々な型および幅でのリサンプリング性能を評価した。一般的な 32bit 単精度浮動小数点と同等のリサンプリング性能を保ちながら、回路面積、スループットが向上できる範囲を求めた。また、3 つの量子化アルゴリズムを提案し、少ないビット数の整数を用いて、32bit 単精度浮動小数点と同等のリサンプリング性能が出る範囲を求めた。