

科学研究費助成事業 研究成果報告書

平成 26 年 6 月 6 日現在

機関番号：62603

研究種目：基盤研究(B)

研究期間：2011～2013

課題番号：23300108

研究課題名(和文) GPU指向の計算集約型統計アルゴリズムの高度化と実用ソフトウェアの開発

研究課題名(英文) Research on advancement of computer-intensive methods for statistical estimations along with Graphical Processing Units and development of application software

研究代表者

樋口 知之 (HIGUCHI, Tomoyuki)

統計数理研究所・所長

研究者番号：70202273

交付決定額(研究期間全体)：(直接経費) 15,200,000円、(間接経費) 4,560,000円

研究成果の概要(和文)：統計計算のアルゴリズムの研究においては、計算機環境の変化にも常に敏感でなければならない。近年、マルチコアやメニーコアと呼ばれる、複数のCPUを格納したマルチプロセッサを搭載した計算機や、画像処理専用の演算装置を汎用計算用演算装置に転用する動きが注目を浴びている。繰り返し計算の多い統計アルゴリズムを高並列計算向きに進化させ、次元の呪いにより従来は困難とされている問題を超並列計算機システムでシンプルに解決することを目標に、多段階層的分散超並列計算機上で効率よく統計的推測を行うための方法を研究した。

研究成果の概要(英文)：A Development of new algorithms for statistical computations needs careful and continuous attention to trends in the computational environments. Recently much interest is laid on making use of the CPU with multiple cores such as the multi-core processor and General Purpose computation on Graphical Processing Units (GPGPU). In this study the statistical algorithms that involve a huge number of a simple iterative calculation are designed to develop in aiming at its implementation on the massive parallel computer. Actually we study the statistical methods which runs on a special computational architecture having the multiple layers, each of which is the distributed parallel computer system. As a result, the massive parallel computation is expected to mitigate a difficult problem arising from the curse of dimension.

研究分野：情報学

科研費の分科・細目：統計科学

キーワード：統計計算 コンピュータ支援統計 ベイズモデリング 並列計算

1. 研究開始当初の背景

統計計算のアルゴリズムの研究においては、計算機環境の変化にも常に敏感でなければならない。実際、統計科学の新しい諸推定手法の誕生には、統計科学コミュニティの内発的な進化よりも、データ環境や計算機性能の激変といった外力が大きく関与してきた。歴史的には、1980年代のいわゆるスーパーコンピュータの計算速度の向上とともに、ブートストラップに代表される計算集約的なアプローチへの統計的推測手法の転換がまず思い出される。同時期に計算機のメモリーも巨大化し、それは巨大な記憶領域を必要とするベイズモデルの実用化に大きく寄与した。その後、ベクトル型スーパーコンピュータがビジネスモデルとして成立しにくくなり、日本の計算機ベンダーの開発能力にも一時の勢いがなくなってきたこと、またインテル等のCPU単体の演算速度の向上も頭打ちになったことなど、HPC(ハイパフォーマンスコンピューティング)分野のイノベーションを生み出してきた計算機環境も新規性に乏しい状態であったのが2000年代前半である。

近年、ビジネスモデルとして成功を取っているのは、マルチコアやメニーコアと呼ばれる、複数のCPUを格納したマルチプロセッサを搭載した計算機である。なおここで、1コアとは一演算器のことで、昔のイメージでは1CPUと理解してもらってかまわない。デスクトップパソコンでも4コアを搭載した市販品が普通になってきている。また、そのような演算装置を複数搭載した、計算サーバの廉価化も著しいものがある。さらには、100を超えるスケールの並列プログラミング環境も実現されつつある。このように、デスクトップパソコンでも並列計算による大きな計算能力を享受できるようになってきた。

これ以外に、近年HPC分野で最も期待を集めているものは、画像処理専用の演算装置—GPU(Graphical Processing Units)と通常呼ばれる—を汎用計算用演算装置に転用する動きである。つまり、GPUをCPUとして活用する潮流である。このようにして利用されるGPUのことを、General Purpose computation on Graphical Processing Unitsの略称で、GPGPUと呼ぶことが多い。元来、GPUは、その名が表す通り、画像処理に特化した限られた演算能力を与えられた演算ユニットにすぎなかったが、そのスーパーコンピュータに匹敵する演算性能から、数値計算に従事する多くの研究者・技術者の注目を集めている。

2. 研究の目的

アルゴリズムの開発においては、計算機環境の変化にも常に注視しておくことが大切である。近年、プロセッサのマルチコア化が劇的に進んでおり、GPGPU(General Purpose computation on Graphical Processing Units)、マルチコア等の100を超えるスケールの並列プログラミング環境もデスクトップパソコンで実現されつつある。逐次モンテカルロ法やブートストラップ、ブースティング等の、計算集約型のアルゴリズムを高並列計算向きに進化させることで、次元の呪いにより従来は困難とされている問題を超並列計算機システムでシンプルに解決することを目指す。その結果、GPGPUから次世代スーパーコンピュータまで、多段階層の分散超並列計算機上で効率よく統計的推測を行うための方法を確立する。

3. 研究の方法

GPGPUの持つ並列計算機としての潜在能力をフルに活用するためには、GPUの設計思想を十分理解した上で、それに適したアルゴリズムの開発が枢要である。そこで、GPGPUの設計に詳しい研究者と、統計科学の計算アルゴリズムに詳しい研究者が共同で研究を行う。GPGPU等の階層的分散超並列システム向きに進化させるアルゴリズムは、粒子フィルタ(逐次モンテカルロ法)、ブートストラップ法、ブースティング等を考えている。使用する計算環境は、既に我々研究チームで保持しているGPGPU及び大規模PCクラスタ計算サーバである。

あわせて、GPGPU以外のHPC環境の動向にも十分留意し、最新のスーパーコンピュータが利用可能な状況にも対応することを念頭にアルゴリズムの開発を進める。開発した計算システムを廉価なノートパソコンに実装することで、金融サービス等のリスク計算の卓上オンライン計算サービス等も今後は可能になると予想する。技術およびビジネスの観点からの実現上の問題点の調査も行う。

4. 研究成果

これまでの準備研究において、素朴な粒子フィルタアルゴリズムのGPGPUへの実装研究を行った。その研究開発においては、a) プロセスエレメント(PE)に直結したオンチップ・メモリーの小容量 b) 大容量である共有メモリーとPE間の通信の著しい低速度の二つが問題として浮き彫りになった。平成23年度は粒子フィルタのアルゴリズムを題材に、このGPGPUの持つ2大弱点の克服に注力した。前者の困難は、粒子フィルタのアルゴリズムというよりも、求解に逐次ベイズフィルタを必要とする一般状態空間

モデル中のシステムモデルの特性（計算サイズ）に由来する問題なので、計算コストやシステムの非線形性と言った、システムモデルの特性に即したアルゴリズムの採用策を探究すればよいことは明らかである。一方、後者は、逐次ベイズフィルタ等の SIR (Sequential Importance Resampling) 法の基本的原理である、リサンプリング（ベイズの定理に基づく更新式の数値的实现）プロセスに関係する問題であるので、根源的解決策が必要である。まず粒子フィルタの分散並列アルゴリズムである、Distributed Particle Filter (DPF) (Ex. Sheng et al., 2005) の近年の研究動向について調査した。その結果我々は、SIS (Sequential Importance Sampling) のような、各時刻で各粒子の計算結果の情報を非同期で扱うアルゴリズムを採用することにした。計算精度を担保した上での計算資源活用効率性を理論的に考察した結果、SIS の計算資源の効率的利用の観点からは無駄である悪い点を勘案しながら、GPGPU の特性に即したアルゴリズムの改良を行った。あわせて、融合粒子フィルタの GPGPU への適用可能性も探った。

粒子フィルタのデータへの適用問題としては、データ同化のパラメータ推定を取り扱う。具体的に言えば、これまでの研究で十分な経験のある、概日周期（動物の行動サイクルがほぼ 24 時間であることを説明する）パスウェイモデルと、遺伝子/タンパク発現データを統合したデータ同化を題材とする。すでに我々は、単体 CPU で長時間計算することで、この問題を解決する方策の有効性を示しているが、GPGPU 上で走る新しいアルゴリズムを利用することにより粒子数の飛躍的増大を実現した。概日周期モデルのデータ同化以外に、アルゴリズム改良実験の題材として適切な問題の発掘調査を行うとともに、将来的に次世代スパコンへの実装を考慮した技術的問題の調査研究を行った。

平成 24 年度は、ブートストラップ法の GPGPU 向きアルゴリズムを研究開発した。GPGPU の潜在的な能力を最大限に活用し、従来の B の値を 100~1 万倍に可能にするとともに、ブートストラップ推定値を精度保証するアルゴリズムを構築できた。本年は、ベイズ統計にかかわる国際会議やそれにあわせたサテライト会議が開催されたので、GPGPU を利用した高効率ベイズ計算や、近年大流行している ABC (Approximate Bayesian computation) を研究している研究者と情報交換を行った。ABC は、計算機リソースを多く要するベイズ統計の中でも、際だってリソースを必要とするもので

ある。同質の繰り返し計算の部分をうまくアクセラレータに埋め込むことが可能なら、ABC の持つポテンシャルは大きく飛躍することは明らかである。このアイデアの実現性について意見交換するとともに調査研究も行った。

最終年度代表者・樋口は、分散超並列指向の線形モデル計算を統合するためのベースアルゴリズムの開発のために、統計パッケージで GPGPU を含めた大規模並列化済みの関数群の調査を行い、開発状況の把握とボトルネックとなっている問題点の同定につとめた。分担者・中野と長尾は、23, 24 年度開発したアルゴリズムを、大規模 PC クラスタ向けに変更する作業を行うとともに、平成 26 年度に稼働予定の本研究所に設置される 3 つのタイプのスーパーコンピュータシステムに実装するための問題点の洗い出しも行った。これまで MPI 等をベースとした並列アルゴリズムの開発を行ってきたが、次期システムの一つは世界最大の共有メモリーをもつユニークなシステムであることから、OpenMP を利用した並列化も視野に入れ、多様な並列化のスキルアップにつとめた。

斎藤は、複数の並列計算システムを利用し、実際のデータに対する判別問題を題材に手法の評価を行った。具体的には、銀河のスペクトルデータから銀河の形状を判別する問題に適用を試みた。その結果を、米国の研究者らがクラウドソーシングにより求めた結果と比較し、科学的な観点から判別結果に大差がないこと、および提案手法が計算効率の面で優れていることを確認した。これまで開発した計算システムを一般ユーザにも利用してもらうため、GPGPU 向けのプログラムを整理し、メモノートを作成した。あわせて、諸オフライン的におこなっている計算サービスの中で、オンライン化により劇的に有益性が増すと考えられるビジネスモデルの調査を行った。

5. 主な発表論文等

[雑誌論文] (計 10 件)

- ① 長尾大道, 樋口知之, 地震音波データ同化システムの開発 — 双子実験による検証 —, 統計数理, Vol. 61, No. 2, 257-270, 2013. 査読有
- ② 長尾大道, 佐藤光三, 樋口知之, マルコフ連鎖モンテカルロ法を利用したトレーサー試験からフラクチャーの物理パラメータを推定する方法, 石油技術協会誌, Vol. 78, No. 2, 197-209, 2013. 査読有
- ③ H. Nagao, T. Higuchi, Data assimilation system for seismoacoustic waves, Proceedings of

- 16th International Conference on Information Fusion, 2013. 査読有
- ④ 長尾大道, 樋口知之, データ同化モデルパラメータの分布推定, 日本機械学会誌, Vol.116, No.1131, 23-26, 2013. 査読無
- ⑤ 本橋永至, 磯崎直樹, 長尾大道, 樋口知之, 事例研究「状態空間モデルによるインターネット広告のクリック率予測」, オペレーションズ・リサーチ, Vol.57, No.10, 574-583, 2012. 査読有
- ⑥ M. Saito, S. Imoto, R. Yamaguchi, H. Sato, H. Sakada, M. Kami, S. Miyano, T. Higuchi, Parallel agent-based simulator for Influenza pandemic, Advanced Agent Technology, Lecture Notes in Computer Science, Springer, Vol.7068, 361-370, 2012. 査読有
- ⑦ M. Saito, S. Imoto, R. Yamaguchi, S. Miyano, T. Higuchi, Identifiability of local transmissibility parameters in agent-based pandemic simulation, Proceedings of 15th International Conference on Information Fusion, 2012. 査読有
- ⑧ S. Nakano, T. Higuchi, Weight adjustment of the particle filter on distributed computing systems, Proceedings of 15th International Conference on Information Fusion, 2012. 査読有
- ⑨ S. Nakano, T. Higuchi, Non-storm irregular variation of the Dst index, Annales Geophysicae, Vol. 30, 153-162, 2012. 査読有
- ⑩ T. Higuchi, Embedding reality in a numerical simulation with data assimilation, Proceedings of 14th International Conference Fusion, 2011. 査読有
- [学会発表] (計 19 件)
- ① 樋口知之, グーグル、マイクロソフトのビジネス展開にみる粒子フィルタ研究のビジネスへの貢献, パーティクルフィルタ研究会 (月例行事第 100 回記念), 2014 年 3 月 28 日, 統計数理研究所 (東京都立川市)
- ② T. Higuchi, Sequential Data assimilation and stream computing for big data, 2014 年 2 月 6 日, 英国大使館 (東京千代田区)
- ③ 樋口知之, データ同化によるエミュレータ・デザイン, 第 18 回先端的データベースと Web 技術動向講演会, 2013 年 12 月 3 日, 国立情報学研究所 (東京千代田区)
- ④ 樋口知之, ビッグデータがせまる人への投資とサービスの高度化, ビッグデータ & オープンデータ研究会 in 九州, 2013 年 12 月 2 日, 福岡 SRP センタービル (福岡市早良市)
- ⑤ 樋口知之, システムのスマート化を支える数理基盤技術, 計測自動制御学会スマートセンシングシステム部会 SICE 計測部門セミナー, 2013 年 9 月 27 日, 東京ビッグサイト (東京都江東区)
- ⑥ 樋口知之, ビッグデータと統計学をつなぐもの, 第 11 回地球シミュレータシンポジウム-さらなる社会・未来・安心への貢献を目指して-, 2013 年 9 月 12 日, 秋葉原コンベンションホール (東京都千代田区)
- ⑦ H. Nagao, Time-series modeling to infer properties of fractures/faults within the Earth's crust, Joint Meeting of the IASC Satellite Conference for the 59th ISI WSC and the 8th Conference of the Asian Regional Section of the IASC, 2013 年 8 月 22 日, Yonsei University (Soul, Korea)
- ⑧ 樋口知之, データ同化は今後の"ものづくり"に欠かせない基盤技術, ビッグデータソリューションセミナー, 2013 年 7 月 17 日, 大崎ゲートシティ (東京都品川区)
- ⑨ 樋口知之, ビッグデータが開く新しい科学像, 日本地球惑星科学連合 2013 年大会, 2013 年 5 月 20 日, 幕張メッセ国際会議場 (千葉市美浜区)
- ⑩ 樋口知之, ビッグデータが拓く新しい情報通信の世界 (~サイバーとリアルの間を取り持つデータ同化技術を中心として~), 電子情報通信学会東京支部講演会, 2012 年 12 月 13 日, 機械振興会館 (東京都)
- ⑪ T. Higuchi, Prediction, Design, and Data assimilation, 2012 International Conference on Modeling, Analysis and Simulation (ICMAS), 2012 年 11 月 9 日, 明治大学駿河台キャンパス紫紺館 (東京都)
- ⑫ T. Higuchi, Data Assimilation: Challenge for Big Data through Numerical Simulation, The Ninth International Conference on Flow Dynamics (ICFD2012), 2012 年 9 月 19 日, Hotel Metropolotan (仙台市)
- ⑬ 斎藤正也, インフルエンザ介入政策における集団免疫を高めるワクチン接種対象群の選択, 2012 年度統計関連学会連合大会, 2012 年 9 月 11 日, 北海道大学高等教育推進機構 (札幌市)
- ⑭ 本橋永至, 時系列モデルを用いたインターネット広告の効果予測, 2012 年度統計関連学会連合大会, 2012 年 9 月 10 日, 北海道大学高等教育推進機構 (札幌市)
- ⑮ S. Nakano, Weight adjustment of the particle filter on distributed computing systems, 15th International Conference on Information Fusion, 2012

年7月12日, Raffles City Convention Centre (Singapore)

- ⑩ M. Saito, Identifiability of local transmissibility parameters in agent-based pandemic simulation, 15th International Conference on Information Fusion, 2012年7月12日, Raffles City Convention Centre (Singapore)
- ⑪ T. Higuchi, Data assimilation and sequential Bayes filters: information fusion with numerical simulation, ISBA 2012 World Meeting, 2012年6月26日, 京都テルサ (京都市)
- ⑫ 樋口知之, アンサンブルベース逐次データ同化と HPC, CPS セミナー, 2012年6月7日, 惑星科学研究センター (神戸市)
- ⑬ 樋口知之, データ同化による半現実世界の実現～オンライン情報統合とアンサンブル予測の計算基盤～, 電子情報通信学会パターン認識とメディア理解研究会 (PRMU 研究会), 2011年10月6日, 幕張メッセ国際会議場 (千葉市美浜区)

[図書] (計 1 件)

- ① 樋口知之, 上野玄太, 中野慎也, 中村和幸, 吉田亮, 朝倉書店, シリーズ<予測と発見の科学>6 データ同化入門—次世代のシミュレーション技術—, 2011.1-15, 31-46, 総ページ数 240

[その他]

<http://www.ism.ac.jp/~higuchi/>

<http://www.ism.ac.jp/>

<http://daweb.ism.ac.jp/>

6. 研究組織

(1) 研究代表者

樋口 知之 (HIGUCHI, Tomoyuki)

統計数理研究所・所長

研究者番号: 70202273

(2) 研究分担者

中野 慎也 (NAKANO, Shinya)

統計数理研究所・モデリング研究系・助教

研究者番号: 40378576

長尾 大道 (NAGAO, Hiromichi)

東京大学・地震研究所・准教授

研究者番号: 80435833

齋藤 正也 (SAITO, Masaya)

統計数理研究所・データ同化研究開発センター・特任助教

研究者番号: 00470047