

サブテーマ3：計算機による帰納的モデリングのための環境

研究代表者

[統計数理研究所] 中野 純司

共同研究者

[統計数理研究所] 田村 義保, 佐藤 整尚, 染谷 博司

[国立情報学研究所] 三浦 謙一, 速水 謙

[国立極地研究所] 岡田 雅樹

[国立遺伝学研究所] 五條堀 孝, 菅原 秀明

1. 研究目標

計算機は現代科学において、欠くことのできない道具である。それは単なる巨大電卓にはとどまらず、科学を記述する形式のひとつになっている。すなわち計算機を用いたモデリングは、伝統的な科学研究における自然言語を用いた記述方式と、数式を用いた記述方式の中間に位置する、第3の記述方式であるといえる。そしてそれは、計算機の巨大な計算能力を利用することにより、非線形で複雑な現象を最も的確に記述できる手段である。

現在では、計算能力があればあるほど、多くの自然現象に対する有用なモデリングが可能であることが認識されており、そのための多くの計算機技術が開発されている。それらのなかでも、スーパーコンピュータ、並列化、ネットワーク、グリッド、乱数、統計解析環境などの技術が特に重要である。

本研究では、立場の異なる研究者が緊密に連絡を取り合うことによって、科学研究に有用な方向にこれらの技術を発展させることを目標とする。具体的には、物理乱数および疑似乱数の発生およびその比較、大規模数値計算と統計解析環境における並列化とネットワークおよびグリッドの利用を研究する。

2. 年次研究計画

研究のコアとして統計解析環境 R を用いる。R はフリーのソフトウェアで、ソースが公開されているため、並列化など必要な改変が可能であり、その設計は汎用的で多くの科学分野で利用可能なものであるからである。特に統計研究においては事実上の標準言語に近いものとなっていることとバイオインフォマティックスのためのシステム Bioconductor の基礎となっていることは重要である。

われわれは R をスーパーコンピュータ上で利用できるようにするとともに、その利用を容易にするような環境を作成する。具体的には、並列計算をできるだけ容易に利用できるような環境、特に Web ブラウザを含むインターネット上からの利用環境用のソフトウェアを段階的に作成する。また、物理乱数、疑似乱数が R から利用できるようにし、その性質の良さも保証できるように研究を行う。Bioconductor のスーパーコンピュータ上での利用を可能にし、遺伝研が作成し

ているデータベースとの協力ができるようにする。さらに、複雑な現象、例えば極地の自然現象なども解析できるような汎用的で強力なシステムを構築する。

なお、関連する技術として、データの可視化も研究していく。それは複雑な現象の可視化はスーパーコンピュータの能力を必要とする場合が多いこと、人間が現象を理解するためにもっとも直観的でわかりやすい手段となれること、のためである。さらに、Rと並んで科学研究でよく利用されているソフトウェアであるMatlabのフリー判と言えるSci labのスーパーコンピュータ上の有効利用も可能になるようにしたいと考えている。それにはRに関する研究の成果が生かせると考えられるからである。

3．平成17年度の研究進捗

まず、スーパーコンピュータ上でRを動かし、並列計算が行えるようにした(一部平成16年度に達成した)。そして、Web ページからその機能が利用できるようなユーザインタフェースおよびサーバプログラムを作成した。これにより、Rの並列的な利用が容易になった。また、並列化R(およびそれを利用するBioconductor システム)を遺伝研のWeb を利用したデータベース利用環境と結びつけるための環境整備を両研究所で行った。また、乱数に関しては、物理乱数公開システムを作成した。さらにRから物理乱数が直接利用できるようにするとともに、情報研で開発されている並列計算機用疑似乱数発生プログラムを利用することを考えた。さらに、極地研での乱数利用技術を移転するためのパイロットシステムの作成に着手した。

4．平成17年度研究成果

(1) 知見・成果物・知的財産権等

- ・Rをスーパーコンピュータ上で動かすための改良(R作成コアチームに連絡済みでRの最新版に随時取り入れられている。)
- ・R実行用Web環境(サーバプログラムおよびクライアントプログラム)
- ・スーパーコンピュータ上での物理乱数、疑似乱数のRでの利用プログラム

(2) 成果発表及び著書執筆等

- ・上記の成果はWeb上で公開する予定である。

5．その他

- ・スーパーコンピュータ技術の発表のためにスーパーコンピュータコンファレンスが毎年アメリカ合衆国で行われている。平成17年度はシアトルで行われたが、そこに参加した。そして、

Rのスーパーコンピュータ上での利用を試みているグループ

Sci lab 開発グループ

との意見交換を行った。また、平成18年度の展示発表のための情報収集を行った。