

1.7 統計的手法の物理解析への応用とシミュレーション

柴田 章博

概要

本プロジェクトでは、ベイズ統計に基づく数値統計やデータマイニングの技術や手法の研究とともに、主として数値シミュレーションの解析や高エネルギー物理や構造解析など加速器実験におけるデータ解析で必要とされるデータ解析法について具体的テーマを取りあげてその解析法を開発・改善を行うことを目的としている。

これまでも、データマイニングや情報処理の手法の応用研究は様々な分野で行なわれている。例えば、高エネルギー実験におけるデータ解析はデータマイニングの過程そのものであり、ニューラルネットワークや自己組織化マップ(SOM)などのデータマイニングの手法が活用されている。近年の高エネルギー実験は、装置の複雑化・巨大化に伴って、解析すべきデータも複雑化・巨大化となるため、複雑な構造を持つデータに対する大規模かつ高速なデータ処理が必要とされる。また、効率の良い実験装置の開発のためにも、シミュレーションと実験開発を解析手法の融合した研究は重要である。

一方、ベイズ統計の枠組みは、データ及びモデルパラメータを確率変数として統一的に扱うことができ、また従来のパラメータ・フィットで使う最尤法を含むため系統的かつ厳密な解析を行うことができるが、一方で、膨大な計算量を必要としていた。近年の計算機の急速な発展と普及に伴い、データマイニング及び情報処理においてベイズ統計に基づく数値統計の研究が急速に発展を遂げ、その成果の様々な分野への応用が行われている。

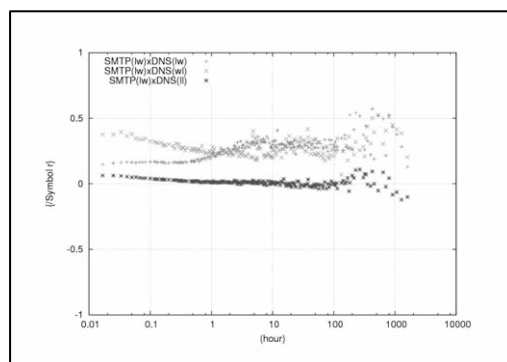
トピックス

時系列データの解析

時系列データを事例として、データ間に潜む関係を見出す試みを行った。インターネット通信は、集中制御の仕組みを持たない自律的な成長と調整の機能を持った時刻動系における流動であり、物理的な解析対象としても興味深い。本研究では、KEKのファイアウォール(FW)の通信のログを利用して、LANとWANの間の通信に潜む関係を検出できるかについて調べた。時系列データの解析には、

DFA(Detrended Fluctuation Analysis)を拡張した、プロトコル別(インターネットサービス別)にトレンドを除去した揺らぎの共分散分析を行うことで、各プロトコルの通信データに潜む揺らぎ起源についての分析を行う。

右図は、電子メールサービスとDNSサービスについてのDFA共分散解析の結果を示す。



各時間スケールにおける相関係数を示している。WAN・LAN間の電子メールの送受信には、アドレス解析の際にDNS参照を行っていることを反映していると考えられる。右表は、各プロトコル間の相関係数の計測結果をまとめたものであり、インターネット上のサーバ間の連携を検知できたことを示している。

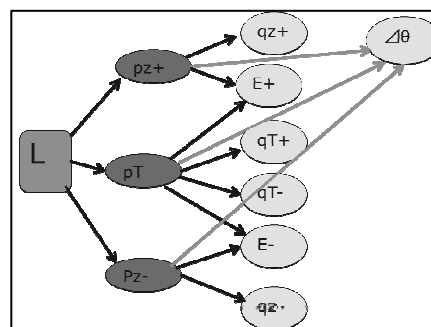
	WEB	SMTP	Others	DNS
WEB	-	×	×	○
SMTP	×	-	×	○
Others	×	×	-	△
DNS	△	△	△	-

表 4: 接続数に対する共相関係数 $\rho_{i,j}$. 表の上三角は LW の通信, 下三角は WL 通信に対する結果を示す. ○は有意な相関が検知されたこと, △は若干の相関が検知されたこと, ×は相関が検知されなかったことをあらわす.

高輝度電子陽電子線形加速器における実験解析を題材として

ベイズ統計の枠組みを用いることで、対象としている変数全体の関係を統一的に扱い、データやモデルパラメータにあるあるデータを対応させ分析する。また、グラフィカルモデリングの手法を用いることで、変数間の関係を視覚的にとらえ、データにある因果律（物理のモデル法則）の考察を見通しよく行うことができる。これまでの研究で、ベイジアン・ネットワークに現れる条件付確率分布関数を連続変数における分布関数に拡張した。これを、高輝度電子・陽電子線形加速器の衝突実験におけるルミノシティ推定の問題に適用し、厳密な粒子反応の統計モデルを構築とベイズ統計に基づく推定を用いることで、従来の研究におけるパラメータ推定よりはるかに少ない（実用的な）データ量でパラメータ推定を行うことができることを示した。また、（加速器パラメータを非対象な設定にするなど）推定するパラメータ数を増やした環境下においても、精度よくパラメータ推定が可能であり、従来提唱の方法を大きく改善するものであることを示した。

今年度は、測定器モデルの検討をおこない、個々の測定誤差が比較的大きな場合におけるパラメータ推定について検討を行った。右図は、測定器モデルのグラフィカル表示である。対象とするパラメータ推定が Bhabha 事象であることを考慮し、計測可能な量についての関係を示している。ベイジアン・ネットワークにおける解析では、計測全体をパラメータ・データ間の関係に基づいて、パートに分解をして考察を進めることができ、パラメータ推定の解析は、Lの部分にルミノシティ推定モデルにつなぐことで実行できる。測定器モデルの検証は、いくつかの測定ができなかった場合（データ欠損）や大きな誤差を含む場合などの条件を変えて分析をおこなうことで、精密測定のシミュレーションを行うことができる。



成果発表

インターネット通信の DFA に基づく共分散分析, 柴田章博, 村上直 第 16 回交通流のシミュレーションシンポジウム 論文集 17-20 (2010)