

平成 21 年 5 月 25 日現在

研究種目：基盤研究 (C)

研究期間：2006～2009

課題番号：18500223

研究課題名 (和文) ベイズ統計によるデータ解析の研究

研究課題名 (英文) Study on data analysis method based on Bayesian statistics

研究代表者

柴田 章博 (SHIBATA AKIHIRO)

大学共同利用機関法人高エネルギー加速器研究機構・計算科学センター・研究機関講師

研究者番号：30290852

研究分野：数値解析、素粒子論

科研費の分科・細目：情報学、統計科学

キーワード：数値統計、ベイズ統計、EM アルゴリズム、多変量解析、データマイニング、高エネルギー実験

1. 研究計画の概要

データマイニングや情報処理において、ベイズ統計に基づく数値統計解析の手法や様々な分野への応用が行われている。高エネルギー実験の解析もデータマイニングの過程そのものであり、蓄積されたデータの中から事象(興味あるデータの組)を何段階かに分けて抽出しデータモデルを解析して物理の現象を理解する。高エネルギー実験は、実験装置の複雑化・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、効率よい開発のためには、実験装置やデータ解析法開発が不可欠である。

本研究では、高エネルギー実験や構造解析などの加速器実験の物理の解析と密接に関係する反応モデルに焦点をあて、データマイニングの統計的アプローチ、特にベイズ統計に基づく方法に焦点をあて、高エネルギー実験のシミュレーションのデータ解析への応用し解析手法の確立を目指す。具体的には、次世代の高輝度電子・陽電子の衝突実験におけるルミノシティの測定問題を題材とする。

ベイズ統計の特徴であるデータとパラメータを統一的に取り扱いや、条件付確率による因子分析をグラフィカルモデルによって分析を行って、従来の解析手法の改善を行うとともに、よりふさわしい方法論の提案を目指す。また、データ解析に必要とされる数値解析や計算アルゴリズムの開発を合わせて行う。

2. 研究の進捗状況

高輝度電子・陽電子線形加速器の衝突実験における新しいルミノシティ推定法を題材に、事象のスペクトラムの推定法を研究した。解析用のデータ生成は、素粒子反応のイベント生成器を活用した。模擬実験(シミュレーション)でデータ生成することで、事象スペクトラムやノイズの厳密なコントロールのもとにモデル評価ができる。

(1) 電子・陽電子弾性散乱の事象を用いたルミノシティ測定モデルについて、グラフィカルモデル(連続変数を含むベイシアンネットワーク)との対応を行うことによって詳細な粒子反応のモデルとして記述した。ビーム生成、反応、測定の各プロセスに分解し因果律を考慮した、厳密な統計モデル構築を行った。

(2) ベイズ統計の特徴であるデータとパラメータを統一的に取り扱いによって、各プロセスに分解し統計モデルの検証を行うことができる。分解したプロセスごとのモデルの構築と評価をイベント生成器と連携して行った。

(3) 効率のよい統計モデルのパラメータ推定法アルゴリズムの検討を行った。近年の本分野の研究成の積極的な導入を検討した。中間層に多数の連続変数を伴うグラフィカルモデルの評価が必要とされる。EMアルゴリズムの適用し、その有用性が確かめられた。

(4) グラフィカルモデルの評価は、多くの計算資源を要求する。厳密な時計モデルの構築と評価のために、学習(パラメータ推定)やモデル評価の数値解析のアルゴリズム開発を合わせておこなった。事象の独立性を利用したMPIによる並列処理のアルゴリズム導入を行った。

(5) 検出器を配置できないなどにより測定データに欠損が生じ、粒子反応の事象を再構成ができない場合や解析対象と酷似する事象が発生する場合について、推定法の検討及び系統的な誤差評価法についても検討を行った。

3. 現在までの達成度

②おおむね順調に進展している。

(理由)

スペクトル関数を多変量分布関数としてあらわすことで、厳密な推定モデルが構築でき従来の方法を大きく改善できた。また並列アルゴリズムの導入によって、効率的なモデルの評価を行うことができた。

4. 今後の研究の推進方策

現実の実験に適用できるスペクトラムの推定モデルとするため、統計モデルの改善とパラメータ推定アルゴリズムの改善を中心に最終年度の研究を行う。

(1) データ欠損や直接観測不能な説明変数を含むような統計モデルに対するパラメータ推定の問題について検討する。解析を行う事象と酷似するデータが混在する場合には、その事象のフィルターリングとパラメータ推定法(モデルセクション)と評価方法が重要である。

(2) 複雑な因子モデルやモデルセクションを伴うベイズ統計による解析には、計算機資源を必要とするが、事象の独立性を利用した並列処理のアルゴリズムや数値解析のアルゴリズムの改善を行う。

5. 代表的な研究成果

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 3 件)

- ① A. Shibata, S. Kawabata, J. Fujimoto, Y. Kurihara, T. Watanabe. "An inference method of luminosity spectrum in future e^+e^- linear collider." Physics letter B645. 12-18 (2007)
査読あり
- ② Akihiro Shibata, "An inference method of luminosity spectrum in a future high luminosity e^+e^- linear collider", Published in PoS(ACAT)048, 2007.
査読あり
- ③ 柴田章博: "EMアルゴリズムによる高輝度電子陽電子加速器におけるルミノシティースペクトラムの推定", Proceedings of 9th Workshop on Information-Based Induction Science. 232-237 (2006)
査読あり

[学会発表] (計 2 件)

- ① 柴田章博: "EMアルゴリズムによる高輝度電子陽電子加速器におけるルミノシティースペクトラムの推定" 第9回情報論的学習理論ワークショップ (IBIS2006)、2006年10月31日、~11月2日(木)、大阪大学中之島センター佐治敬三メモリアルホール
- ② Akihiro Shibata, "An inference method of luminosity spectrum in a future high luminosity e^+e^- linear collider", XI International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT2007) April 23-27, 2007 at Nikhef, Amsterdam, the Netherlands