

氏名	藤井 恵介
所属機関	京都大学 工学研究科
研究題目	データ駆動型キャリブレーション法の実現による科学計測システムの超高精度化

1. 研究の目的

背景と目的

新たな物理現象の発見、解明には高精度な計測装置が必要不可欠である。しかし一方で、計測システムが高度化するにつれ、そのデータの解釈も難しくなってきた。特に、計測システムにキャリブレーション誤差があると、システムから生み出される計測データにもその誤差が蓄積される。この誤差はシステムノイズや系統的誤差と呼ばれ、ランダムに信号が乱されるランダムノイズとは区別される。

これまで用いられてきた古典的な統計学では、このようにランダムに揺らがないシステムノイズの取り扱いができなかった。本研究の目的は、情報学の分野で発展してきたベイズ統計を用い、大量の計測データからこのシステムノイズを定量評価し、較正する手法を開発することである。

開発する手法

システムノイズはランダムに揺らがない、複数の実験データで類似な成分である。本研究では、図 3 に示すように大量の多チャンネル計測データから類似部分・固有部分を分離し、さらに類似部分をシステムノイズと共通信号に、固有部分をランダムノイズと固有信号に分離するアルゴリズムを開発する。システムノイズは実験間で揺らがないため、一度評価することができれば較正に用いない他の計測データについても補正することが可能となる。

2. 研究の内容(手法、経過、評価など)

本研究では、大量の計測データからランダムノイズ・系統的ノイズの特徴を抽出する方法を開発する。特に、これまでよく仮定されるような一定分散のガウス分布でないようなノイズがある際の手法の開発に取り組む。

具体的には、核融合研究に用いられるトムソン散乱装置による計測データを例にとる。トムソン散乱は非線形信号処理を用いるため、そのノイズ分布はガウス分布とは大きく異なることが予想される。また、大規模な計測システムであるため、詳細な較正も難しいため、系統的誤差も多分に含まれる。一方で、同じシステムで大量のデータが計測されていることも特徴である。典型的な計測信号を図 1(c)に示す。

信号から、真の値、ランダム・ノイズ、系統的ノイズを補正するため、信号の確率分布をモデル化する。ただしその分布が未知であることから、十分に柔軟なモデルを用意し、大量の計測データからその分布を決定する。具体的には、真の値、ランダム・ノイズ、系統的ノイズをニューラルネットワークを用いてモデル化した。真の値の分布は、正規分布に従う確率変数 z とニューラルネットワークによる変換 \mathcal{F} を用いて、 $f(R) = \mathcal{F}(z, R)$ とモデル化した。また、非一様・非ガウシアンノイズについては、 $p(n|f) = St(n|\mu(f), \sigma(f))$ のように、平均 $\mu(f)$ 、分散 $\sigma(f)$ の学生t分布に従うとモデル化した。 $\mu(f)$ 、 $\sigma(f)$ も同様にニューラルネットワークである。最後に系統的ノイズについても、平均0の学生t分布に従う確率変数 m を用いて、 $(e^m - 1)f$ というように、真の値 f に比例するとしてモデル化した。

2. 研究の内容(続き)(書ききれない場合には、同一形態のページを追加しても結構です)

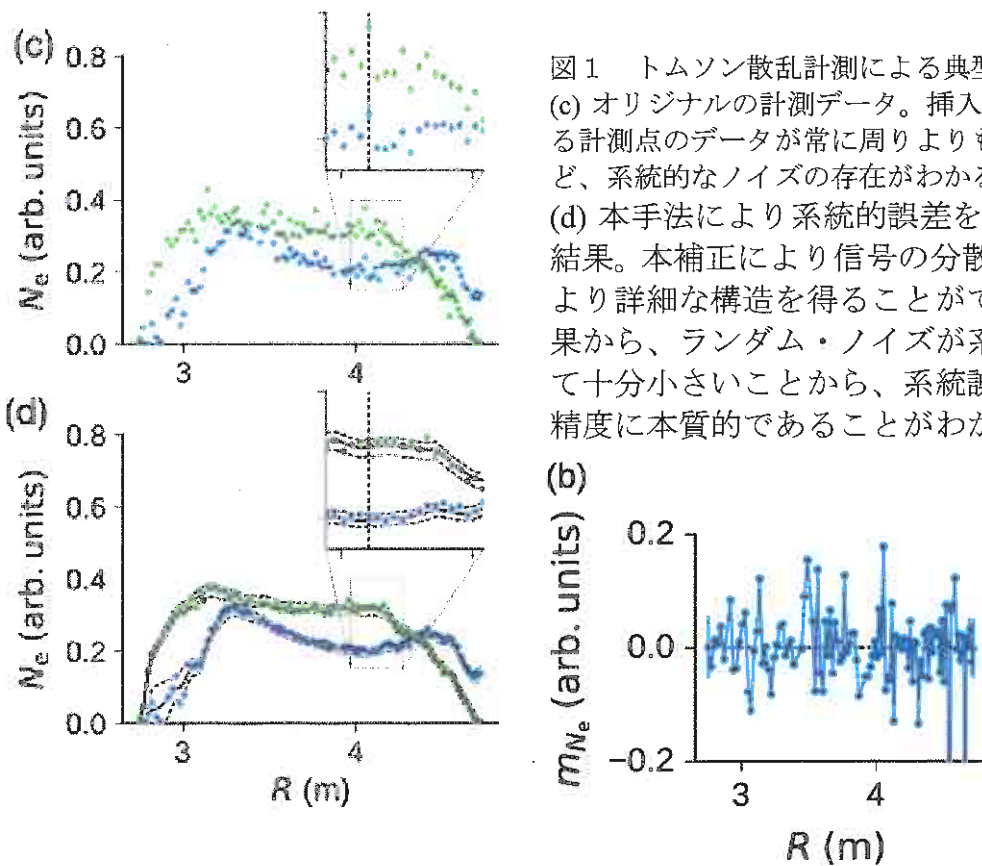


図1 トムソン散乱計測による典型的な計測データ。
 (c) オリジナルの計測データ。挿入図に示すように、ある計測点のデータが常に周りよりも高い値を取るなど、系統的なノイズの存在がわかる。
 (d) 本手法により系統的誤差を推定し、補正した結果。本補正により信号の分散が大きく減少し、より詳細な構造を得ることができた。また本結果から、ランダム・ノイズが系統的誤差に比べて十分小さいことから、系統誤差の補正が計測精度に本質的であることがわかる。

行。

これらの未知係数をデータから推定するため、約 10^6 セットのデータ y を用意した。これらのデータからニューラルネットワーク $p(y|\Theta) = \int p(y|z, m, \Theta)p(z)p(m|\Theta)^{M/\alpha} dz dm$ を最適化する。

本手法により評価した系統誤差の値を図2に示す。また、求めた値を用いて元の計測信号を補正したものを図1(d)に示す。ほとんどの信号のばらつきが補正された。ここから、解析したトムソン散乱計測システムでは、元の信号に含まれるノイズのうち90%程度が系統誤差に由来すること、それを補正することで10倍以上の計測精度の向上が実現していることが明らかになった。ランダム誤差の大きさを図1に点線で示す。信号強度の小さい領域(大半径 $R \sim 3$ m)では分散が大きいたことが予想された。これは、トムソン散乱の計測原理と合致する。

3. 研究の結論、今後の課題

本研究により、非一様非ガウシアンノイズ、系統的ノイズがあるときのデータ解析法を開発することができた。本手法を、核融合研究のためのトムソン散乱計測システムに適用した。本システムの場合、系統的ノイズが主なノイズ源であったため、その補正により当該システムの計測精度を 10 倍に向上することが可能となった。

ただし同一の計測システムによる大量のデータが存在すること、さらにデータの分布、計測システムの構成が同一であることが、本手法の前提となる。しかし実際には、計測システムやデータの分布が時間的に変化することも一般的である。現状のモデルでその状況に対応するためには、モデルの学習を継続的に続けることが必要である。将来本手法を実践的に用いるには、データの分布の変化・システムの些細な構成の変化により頑健なモデルが求められる。

4. 成果の価値(とくに判りやすく書いて下さい)

4. 1. 社会的価値

2016 年始めに発表された重力波の史上初検出が非常に大きなニュースになったように、科学は「計測の高度化」によって駆動されている。

本研究が完成すれば、多くの分野で計測システムの高精度化が実現される。これにより、広い学術分野でこれまで捉えきれなかった計測データが得られると期待される。特に、申請者の所属する核融合分野で新たな知見を得る事ができれば、より早期の核融合発電実現に貢献できると考えられる。

4. 2. 学術的価値

これまで科学計測システムは、別途行うキャリブレーション実験により較正されて用いられてきた。しかし原理的な制約のためキャリブレーション精度には限界があり、それが計測精度自体を制限してきた。本研究が完成すれば、計測データを用いることで計測システムのキャリブレーション精度の向上が可能となる。

科学計測システムの較正という概念が変革されるとともに、計測精度向上により広い科学計測分野において新たな知見が得られることが期待される

4. 3. 成果論文(本研究で得られた論文等を年代順に書いて下さい。未発表のものは公表予定を書いて下さい)

[国際学会発表]

” Machine Learning of Noise in LHD Thomson Scattering System”

2nd IAEA Technical meeting on Fusion Data Processing, Validation and Analysis

Keisuke Fujii

May 30th – June 2nd, 2017

[査読付論文]

Robust Regression for Automatic Fusion Plasma Analysis based on Generative Modeling

Keisuke Fujii

Plasma Physics and Confinement Fusion に提出する予定(2018年9月)