

「放射光計測と情報科学・統計数理の融合 ～計測限界突破と新しい知の獲得への道～」企画説明

水牧仁一郎（熊本大理）、上野 剛（理化学研究所）

近年、機械学習の技術が急速に発展しており、その技術をサイエンス分野に適用する動きが活発化しています。放射光計測データに対しても同様で、将来にわたり必要不可欠な技術となることは間違いないと考えられます。機械学習は Fig. 1 に示すように放射光実験を行う際にたてる「計画」、また「計測」そのものに対しても適用可能な技術ですが本特集では「データ解析」に焦点を当てています。「データ解析」は分野・測定手法を問わずに測定対象を理解するのに必要な手段で、測定手法と同等に重要なプロセスです。その「データ解析」において、新しい視点からのデータ整理により新しい科学的概念の創出が期待されています。そこで本号では、「放射光計測と情報科学・統計数理の融合」をテーマに特集を企画し、4編の記事にて構成しています。

情報科学・統計数理において発展してきたベイズ推定、スパースモデリングやビッグデータを取り扱うための階層的クラスタリングなどを用いた事例について扱っています。ベイズ推定はベイズ統計学に基づく統計学的推定手法であり、分光スペクトルや回折・散乱プロファイルにおけるピークの位置・形状・強度などのパラメータの推定に用いられているだけでなく、近年ではピーク生成の起源をモデル化（例えば X 線光電子分光素過程のハミルトニアンをモデル化）することにより、データに含まれるより深い情報を抽出することにも利用されています。またスパースモデリングは、ノイズが存在するデータにおいて起こる過適合（オーバーフィッティング）を防ぐための正則化を行うための手法の一つであり、L1 正則化とも呼ばれています。データを説明するための変数がスパース（疎）であることを利用しています。この2つの手法が一つ一つのデータに対して有効なものであるのに対して階層的クラスタリングはビッグデータ等の大量のデータを対象とする情報科学的な分類・分析手法となります。各記事において、放射光を用いた測定データに対してこれらの手法を用いた解析法の有用な点を各先生方に述べていただいております。

1つ目の記事は、「ベイズ計測を用いた仮想計測解析（VMA）による実験条件設定」と題して、片上舜氏（東京大）にご執筆いただきました。ベイズ統計に基づき、計測データから対象の系の物理モデルを抽出し、それらのパラメータを推定する方法論について、中性子非弾性散乱を

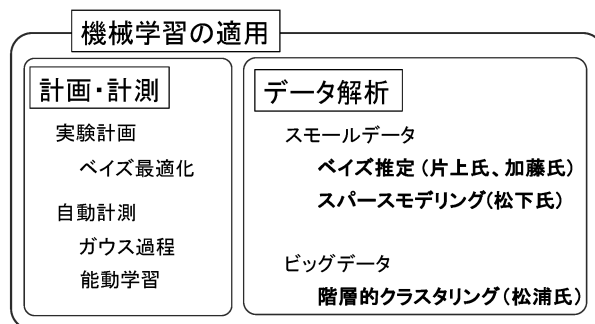


Fig. 1 放射光測定における機械学習の適用

用いたフォノンの分散関係の推定を例にわかりやすく説明いただきました。また放射光分野の実験研究者がデータから様々なパラメータを推定するプロセスにおいて直面する問題点とその解決策を示していただいております。

2つ目の記事は、「データ駆動型全散乱法の開発と応用」と題して、加藤健一氏（理化学研究所）にご執筆いただきました。全散乱法において測定の時に注意を払わねばならない点を説明いただいた後、データ解析においても系統誤差や打ち切り誤差を克服する方法を平易にかつ明瞭に記述されており、全散乱法における問題点の解決が平均化されたものではない不均質構造の理解へ繋がることを示していただいております。

3つ目の記事は、「L1 正則化を用いた光電子ホログラフィーによる原子配列解析」と題して、松下智裕氏（奈良先端大）にご執筆いただきました。光電子ホログラフィー測定の基礎的な概念から測定装置の原理まで詳細に記述いただき、測定データの前処理・補正における問題点を明瞭に記述していただきました。また L1 正則化を用いて測定データから立体原子像を再構成する方法を平易にご説明いただき、この解析法により欠陥を含んだ原子配列の抽出が可能となることを示していただいております。

最後の記事は、「タンパク質 X 線結晶構造解析におけるビッグデータ化と構造生物学における新たな展開」と題して、松浦晃明氏（理化学研究所）にご執筆いただきました。SPring-8・BL32XU で開発されてきたタンパク質 X 線結晶構造解析の自動化と、その技術を生かした高データレート化から生じる大量データへの階層的クラスタリング適用

による結晶構造の分類についてご説明いただきました。この方法により解析データの高精度化のみならず構造多型の有無を判断するのに強力な手法であることを示していただいています。

本特集号の記事によって、読者の方々がデータ解析に情報科学の手法を取り入れてみたいと思う契機になれば望外の喜びです。