

放射光実験におけるオンサイト機械学習の活用と課題

○沓掛健太郎

理化学研究所革新知能統合研究センター

On-site machine learning in synchrotron radiation experiments

○Kentarō Kutsukake

¹ Center for Advanced Intelligence Project, RIKEN

近年、実験・計測機器の性能向上に合わせて、得られる計測データの量は飛躍的に増加しており、特に、大強度ビームによる高速・短時間の測定が可能な放射光実験では顕著である。しかし、計測データの複雑化・高精細化・大規模化・多数化に伴い、データの解析にかかる時間・コストも飛躍的に増加している。解析のために、人手による細やかな条件の設定が必要になる一方で、数千・数万のデータを手作業で処理することはもはや不可能であり、合理的な基準・手法による自動処理が必須である。大型施設での実験のもう一つの課題は、マシンタイムの制限である。限られた実験時間の中で、目的の条件・位置（信号強度が大きい、特定の結晶構造・組成・配向を持つ、など）を探す、いわゆる「条件出し」と呼ばれる探索作業に要する時間は少なくない。また、数か月後の実験機会ではなく、得られたデータをその場で直ちに解析し、その結果に基づいて次の実験を行うことができれば、研究速度は大幅に向上できるであろう。

このような放射光実験における諸々の課題に対して、我々は機械学習を活用したアプローチを検討している。機械学習は、コンピュータを用いてデータから有用な法則を導く方法であるが、近年は様々な分野で応用が広がっており、計測分野においても、計測インフォマティクスとして、目的信号の抽出やノイズの除去など、顕著な成果が報告されている。放射光実験への応用として、我々は特に、オンサイトでの機械学習に取り組んでいる。測定したデータを用いてオンサイトで機械学習モデルを作成し、さらにそのモデルを活用した解析によりその場で次の条件を決定し、次の測定を行うことを企図した。これにより、大量データ解析、迅速な条件出し、研究速度向上の実現を目指した。

図1に今回行った放射光実験（ μ ビーム X線回折）の模式図と X線回折パターン例を示す。試料には表面内で組成と結晶配向に微小な揺らぎを持つ SiGe 薄膜

を用い、試料に照射した μm 径の X 線ビームからの回折パターンを 2 次元検出器で計測した。さらに試料を x - y 方向に移動させることで、試料表面内の各点からの回折パターンを取得した。したがって、ここで得られるデータは 2 次元パターン \times 2 次元座標 = 4 次元の高次元データである。従来は、回折パターンに対してピークフィッティングすることで、結晶の組成と配向情報が解析されてきたが、人手によるチューニングが必要であることや、複数のピークを持つような複雑なパターンに対しては解析が難しいことなどから、大量データの迅速な解析は困難であった。

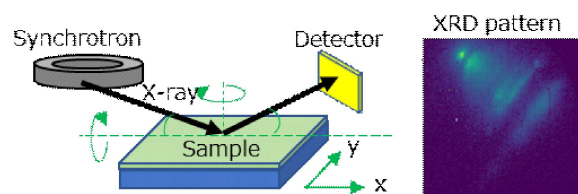


Fig. 1. (左)放射光実験系、(右)X線回折パターン

オンサイト機械学習には、機械学習応用への一般的な要件（目的に対する適切な表現力や説明力）に加えて、モデルの学習と推定が高速であることが求められる。本研究では、教師なし機械学習手法の一つである非負値行列因子分解を用いた。この方法は、大量かつ多様な回折パターンに共通する基底パターンを求め、各パターンを少数の基底パターンの線形和で表す。本研究では、得られた各基底パターンを X 線回折についての専門知識によって意味づけすることで、基底パターンの係数から組成と結晶配向の情報を取得した。ここでの推定は高速に行われるため、試料表面でのマッピング測定後ただちに組成と配向の分布情報を得ることが可能である。さらに、得られた分布から特徴的な組成や配向を持つ領域に対して、より詳細な X 線回折測定を行うことで、局所的な構造を明らかにすることができた。講演では、オンサイト機械学習の課題も含めて、具体的な研究事例の紹介とその可能性の議論を行う。

*E-mail: kentaro.kutsukake@riken.jp