データ駆動科学の物質科学への展開

Data-driven Scientific Approach to Materials Science

福島 孝治 (東大総合文化)
Koji Hukushima (Univ. of Tokyo)
E-mail: hukusima@phys.c.u-tokyo.ac.jp

深層学習の登場以来，昨今の機械学習の発展は社会に対して大きなインパクトを与えているが，科学技術の分野もその活躍が大いに期待されている．自然科学の方法論を概観してみると，まず経験科学の発展に伴い「実験・計測」の方法が整備され，数理科学の発展による「理論」の方法が整理されてきた．これらは第一の方法と第二の方法と呼ばれており，長い間，二つの方法，すなわち「理論」と「実験」による知見のループを成熟させることにより自然科学は大きく発展していた．1980 年代になると計算機の大幅な発展により「シミュレーション」の方法が認知されるようになり，実験困難な状況の理解や理論的に複雑な問題へのアプローチとして独自の役割を担うようになってきた．これが第三の方法である．3 つの方法によるループが科学の発展の加速を可能とした．しかしながら，より最近になり，巨大計算機による大規模シミュレーションや大型施設による高精度実験が可能となると，取得されるデータは巨大となり，そこから有益な情報を引き出すことが難しくなってきている．この困難のためにループが実質的に回らないことが問題として顕在化するようになる．

そこで，ビッグデータを取り扱うデータ科学が第四の方法として期待されることになる．実験でもシミュレーションでも得られたデータが存在するのであれば，情報抽出を目的とするデータ科学の必然性は高まり，さらに機械学習の技術の発展が大きな役割を担うことが期待されるわけである．実際に機械学習の目的の一つである「予測」を実践する場として，物質探索への応用が始まっている．例えば，既存のデータから特定の機能をもつ新物質予測が機械学習により高速化され，これまでの研究者による探索よりも速く新機能物質を見つける例が出てきている．ただ，これでは科学の方法論としては普遍的な第四の方法と呼ぶ必要があるのか個人的には疑問が残る.「シミュレーション」や「実験」の解析手法としてデータ科学の手法が取り入れられたことと質的な違いがどこにあるだろうか.

データ科学の方法論の中での関係性としてありうるもう一つの可能性は，3 つの方法の潤滑剤としての役割を担うことが考えられる．例えば，実験によるデータから情報抽出をし，次の実験計画に指針を与えたり，あるべき理論に条件を提示することである．予測よりも理解を重視した考え方である．大量のデータの中から人では見出すことができず，見逃していた情報を機械学習により取り出し，理解のためのヒントを探ることである．これは「データの科学」というよりは「データ駆動科学」と呼ぶべきであろう．名前の呼び方は重要ではないが，概念を成熟させるとともに，ある分野に限定的な方法ではなく，分野に依らない普遍的な方法論が見いだせるかどうかは研究の展開を見定める必要がある.

講演では，データ駆動科学の実践例を物質科学の例を参照にしながら，プラズマ科学も含めた自然科学での問題点を議論したい.

© 2018年 応用物理学会