

# インフォマティクス技術を実現する次世代高分子材料設計ソフトウェアシステム

(国研) 産業技術総合研究所

機能材料コンピューショナルデザイン研究センター

(兼務) 産総研—東北大数理先端材料モデリングオープンイノベーションラボラトリ

森田裕史

E-mail h.morita@aist.go.jp

## 1. はじめに

近年、マテリアルインフォマティクス技術について、多くから興味もたれており、ある意味ブームになっているといえる。これは、2011年に米国オバマ大統領が Material Genome Initiative (MGI) を発せられることではじまっているが、それ以降米国発で情報を使った材料開発技術について研究が進んでいる。一方、国内もそれに遅れて 2015 年から動きがあり、JST のプロジェクトとして、情報統合型物質・材料開発イニシアティブ(MI2I)が物質・材料研究機構(NIMS)を中心に立ち上がり、2016年には、産総研を中心として NEDO プロジェクトの超先端材料超高速開発基盤技術プロジェクト(超超プロジェクト)がスタートした。一方で、これらとは別の流れではじまった数学、特に幾何学との連携における材料開発の研究について東北大学 AIMR を中心に進められているが、産総研はこの数学を用いた材料研究を進めている東北大学 AIMR と連携し、産総研・東北大数理先端材料モデリングオープンイノベーションラボラトリ(MathAM-OIL)を東北大学片平キャンパスに立ち上げ、「数理マテリアルインフォマティクス」を掲げた研究が進行している。

このような動向に対して、このようなインフォマティクス技術を用いた研究を進めるためには、データを扱うためのソフトウェアが必要となる。我々は従来よりソフトマテリアル統合シミュレーションシステム OCTA を開発してきているが、その OCTA もこのブームにたがわず、情報科学を取り入れたソフトウェア(ツール)としての対応を進めた OCTA8.2β を 2016 年にリリースした。本稿では、我々が開発してきている OCTA を軸に、ソフトウェアの視点からインフォマティクス技術について展望する。さらに我々が関わっている MathAM-OIL などの研究プロジェクトについても紹介し、数学を用いた高分子材料研究の可能性についても議論する。

## 2. OCTA システムの展開

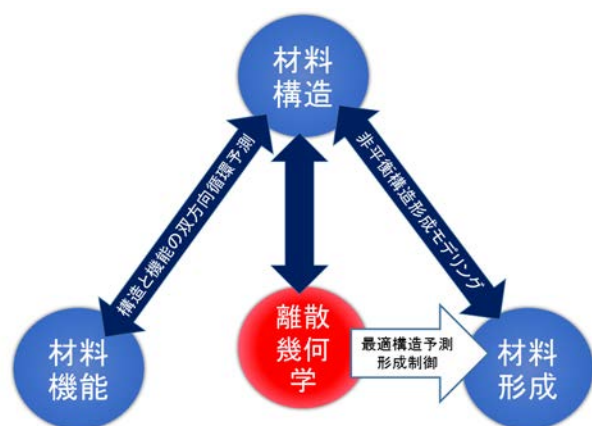
OCTA システムは土井正男らによって開発され、2002年に初めのバージョンがリリースされた高分子材料シミュレーションシステムである。2002年以降、10回以上のバージョンアップをボランティアの手によってバージョンアップを重ねながら、現在に至っており、フリーで配布されてきた。OCTA の主な特徴として、UDF(User Definable File format)ファイルによって様々なシミュレータによるデータを OCTA のプラットフォームを介してやり取りを行えることと、Python を用いてこの UDF ファイルの PrePost の様々な処理を行うことができること、この2つがあげられる。この2つの点を利用することにより、様々なシミュレータの入出力ファイルのデータをやり取りするだけでなく、実験のデータや画像の取り込みも行うことができている。画像の取り込みについては、2003年からの JST-CREST プロジェクトにおいて、その技術が開発されている。

一方で、2016年にリリースした OCTA8.2β について説明する。深層学習・機械学習を扱うためのソ

フトウェアの多くは、実は Python で書かれており、Python なしにこれらをコントロールすることは難しいといっても、過言ではない。OCTA は、元来シミュレーションの結果の描画やデータ解析について、Python プログラムで行っている。このことから、OCTA は Python と親和性がもともと良い。よって、OCTA8.2βには、機械学習に対応した Python ライブラリを同梱し、そのライブラリを用いて機械学習に適用できるようにしている。具体的には、SciPy, Scikit-learn のライブラリを OCTA に持たせている。実際に動作できる証拠として、機械学習のサンプルとしてよく用いられる Iris データベースを用いた機械学習のサンプルを OCTA で実行できている。また、画像の取り込みについても、以前の CREST プロジェクトにおいて開発した技術を用いて行っており、この OCTA8.2βで実行できることも、確認できている。以上より、画像を用いた機械学習解析も行うことが可能であることを示している。

### 3. 数理マテリアルインフォマティクス

産総研では、様々な大学との連携を目的にした拠点として、様々な大学にオープンバージョンラボラトリーを設置している。一方で、東北大学では、2017年3月まで WPI-AIMR を組織し、離散幾何学を材料設計に役立てるための研究が行われてきた。この両者が 2016 年に結び付き、産総研—東北大学 AIMR の連携を構築するために、はじめにでも紹介した MathAM-OIL が立ちあがった。



MathAM-OIL では、材料機能、材料構造、材料形成の3つに着目し、その3つの間を繋ぐための「構造と機能の双方向循環予測」及び「非平衡構造形成モデリング」の研究を行い、さらに、その材料構造に対して、離散幾何学を元に記述子を考えながら、最適構造予測・構造形成制御に繋げる研究を進めている。この MathAM-OIL では、材料に制限をつけて研究を行っておらず、高分子に限らず、広く材料の研究を進めようと考えている。

このような枠組みで高分子材料を考える際には、材料形成が非常に重要となる。高分子材料による製品には、生成した材料が非平衡状態のままになっている場合も少なくない。特に、その形成過程においては、高次構造形成過程が重要であり、多階層性を持っていることが多々ある。このような複雑な構造の理解のためには、その構造を単純化して理解する手法として、離散幾何学は威力を発揮できるツールであると、我々は認識しており、その適用を前向きに考えている。この離散幾何学の検討については、機械学習や深層学習等を用いた研究における構造における記述子の問題として、重要になるのではないかと考えている。通常、機械学習や深層学習における記述子には、物理的な意味を持たない場合が多々ある。しかしながら、離散幾何学を用いた解析結果を経由させて、構造を定義することができれば、構造と機能の相関にも意味を持たせる可能性がないかと考えている。

離散幾何学を用いた解析の1つとして、ホモロジー解析があげられる。東北大学 AIMR の小谷、平岡らは、ランダムな無秩序構造の中に存在する部分的な秩序構造を見出す手法の1つとして、Persistent Homology Diagram (PD)法を用いた解析を提案している。数学的に0次元、1次元といった構造を長さ(大きさ)をパラメータに定義し、その大きさ毎にヒストグラムで表した解析が PD 法である。この解析法の成功例として、ランダムな SiO によるガラスの構造内に、特徴的な秩序構造を見出し、その

構造が機能と関係がある例を示している。現在高分子等のソフトマテリアルに対して適用を考えており、今後検討を進める予定である。

#### 4. 今後の課題

今後、機械学習・深層学習を用いた材料研究、及び数学を用いた材料研究、は劇的に進むと思われる。これらの研究において、材料において機械学習・深層学習との相性もあると思われ、非常に進む研究領域と時間のかかる研究領域が出てくると思われる。この高分子材料に関係する研究については、解析への適用が進むと思われるが、材料自体をピンポイントで示すような研究にまでは、まだまだ時間がかかると考えている。その理由は、機械学習・深層学習を行う際の学習データをどれくらいの範囲までパラメータとして持っておくのかということが、まだまだ理解に時間がかかると思われるためである。ゴムなどは、同じ材料、同じレシピ、同じ装置、同じ条件で材料を作成しても、日によって、人によって全く物性が異なるものができることがある。日や人によって変わる要因がどこにあるのかについて理解が必要であり、それらを機械に学習データとして入れる必要があると思われる。

次の問題が、この学習データの集め方である。深層学習を行うためには、100~1000個レベルの多くのデータがパラメータ付きで、しかも同じクオリティレベルのデータとして必要となる。これらをどのように集めるのかということは、今後も課題となる。実験を通じて構造データを用意すると、顕微鏡で100枚を超える画像をとると、そのサンプル作成から測定までその労力は膨大となる。またエラーを含むデータで学習させた際に、その点がデータ点の少ない領域の点であった際には、学習にも影響が及ぶことが考えられる。よって、学習データ作成は重要な課題である。

もう1つ大きな問題は、知財の問題である。機械学習に関係する知的財産管理についての法整備は、2016年に議論が始まったばかりという状態である。特に、学習済みAIをどのように扱うのか、さらには学習済みAIが生み出したものの知財の扱いも未決である。これらの課題について、今後の動向をウオッチする必要があると思われる。