



— 材料探索で計算と実験を協働・連携 —

多数の記述子を用いて物性との相関を解析／コニカミノルタ様

コニカミノルタ株式会社は、材料研究に計算科学やデータサイエンスを活用しており、近年ではマテリアルズ・インフォマティクス(MI)を社内に普及させることに力を入れています。とくに、計算グループと実験グループとの協働・連携が研究開発の特徴となっており、実験研究者が自ら計算科学ツールを活用するスタイルも定着しています。その中で日常的に重要な位置づけで利用されているのが、イタリア Alvascience 社の「alvaDesc」。「これがないと、もう仕事にならない」とまでご評価いただいています。今回は、開発統括本部 要素技術開発センター DS 技術開発室 解析ソリューショングループ 課長の押山智寛さん、解析ソリューショングループ アシスタントマネジャーの奥山倫弘さん、同じく要素技術開発センター コア技術開発室 コア材料開発グループ アシスタントマネジャーの中澤幸仁さんに活用方法などを聞きました。

—最初に会社紹介とお仕事内容の紹介をお願いします。

押山さん：当社は、売り上げの8割を海外で占めるB2B企業です。ご存じのように、カメラ／写真用フィルムで創業しましたが、材料、光学、画像、微細加工などの技術を強みに、現在はデジタルワークプレイス事業、プロフェッショナルプリント事業、ヘルスケア事業、インダストリー事業の4本柱で事業展開しています。とくに、コア技術とデジタルトランスフォーメーション(DX)の力で、お客さまのさまざまな「みたい」に応えることをミッションにしており、がんの種類や状態、薬の効き具合をナノ蛍光粒子で“診る”「Quanticell」、新しいX線デバイスでこれまでに見えなかった軟骨のようなものまで“視る”「タルボ」、老人介護の見守りのようにAI予測で人の動きを“観る”「HitomeQ」などの技術を提供しています。

われわれの仕事に関係したところというと、コニカミノルタの材料開発は、サイエンスとデータサイエンスの融合による新しい価値の創造を目指しています。従来からの研究者の資質である勘と経験や論理的・演繹的な思考に加えて、データサイエンスを武器として活用することにより、大幅な開発加速や非連続的な価値向上といった効果を引き出すことができると考えています。4～5年前からは、社内教育だけでなく博士人材を採用することにも力を入れ、計算科学やデータサイエンスに強い人を入れるなどして全体の底上げを図ってきています。有識者のドメイン知識や計算科学によるシミュレーションを通じた演繹的な推論、マテリアルズ・インフォマティクス(MI)やプロセス・インフォマティクス(PI)によるデータ駆動型の帰納的な推論、さらに実験系の独自の分析技術を組み合わせることで本質的な理解を進め、課題解決につなげていこうという戦略です。

中澤さんのところは、要素技術開発センターの中でも、計算主体のわれわれとは組織が違って、主に実験でいい材料を見つけたり、メカニズムを探ったりするようなことを行っておられます。役割を違えつつ連携して動いているという感じです。

データ科学を適用、帰納的アプローチと統合

—そうですね。計算やインフォマティクスの方と実験の方が一緒にお仕事されているんですね。

中澤さん：そうですね。われわれの方も完全に演繹的というよりは、データサイエンスの帰納的なアプローチも混ぜ合わせてシフトしていているという実態があります。

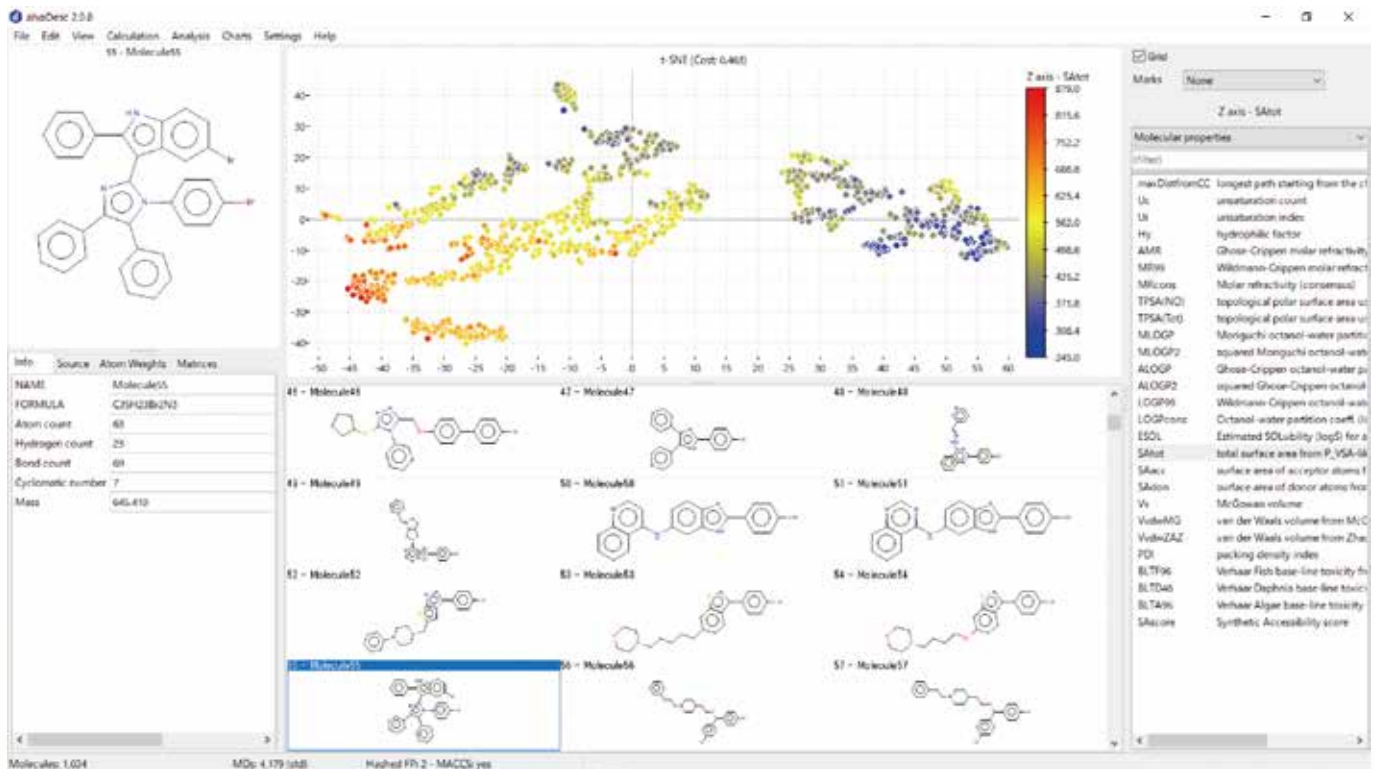
奥山さん：4年ぐらい前からデータサイエンスで良い結果が出るようになってきて、それが実際の開発プロジェクトに益になることがだんだんわかってきました。そういうことがあって、3年ほど前から、デジタルプラットフォーム化が具体化してきました。実験グループの中からデータサイエンスを評価してくれる人も出てきましたので、いまは積極的に社内に広めて、データサイエンスを根付かせようとしています。

—わかりました。では、「alvaDesc」についてですが、最初はこのソフトの前身に当たる「DRAGON」をご導入いただきましたが、その際の経緯をお聞かせください。

奥山さん：導入に当たり最も評価したのは、化合物に対する記述子が豊富に用意されていることです。いまだと5,000個以上の記述子を計算できます。「DRAGON」がないころは、



コニカミノルタでは材料研究でMIなどの活用が進む



alvaDesc GUI 操作画面

化合物の特徴量を自分で探してこないといけませんでした。しらみつぶしに探してみても、20個か30個が限界なんですけど、必死に見つけたデータで回帰分析し、予測モデルをつくってみても、その特徴量と性能の相関がなかなか得られません。記述子がたくさんあれば良い相関が得られるのではないかと考えていた時に、ご指導いただいていた東京大学の船津公人先生（現在は奈良先端科学技術大学院大学特任教授、東京大学名誉教授）から「DRAGON」を教えてくださいました。「DRAGON」を使うとうまく相関がとれ、複雑な性能の予測に使えるということで、約5年前に購入し、活用させていただいています。

特徴量を求めることは第一原理計算でも可能なのですが、1個1個の化合物に対する計算がものすごく時間がかかることと、データサイエンスで取り扱う化合物は数万単位になりますから、トータルで膨大な時間になってしまいます。「DRAGON」(alvaDesc)なら数分で計算できてしまいますから効率的です。

豊富な記述子が魅力、なくてはならない存在に

—それでは、alvaDescをどのようにお使いでしょうか。

中澤さん：わたしの主な仕事は複合材料の研究ですが、樹脂と混ぜる低分子の探索に「alvaDesc」を利用しています。ある物性について、高くなるようにとか、特定の範囲に入るようにとかいうかたちで探索を実施します。最初、100くらいのデータを集めて予測したところ、絶対値はそれほど合わなかったのですが、所望の温度以上とか以下とか、閾値を判定するところでは使えるなと思いました。その後、別の物性を検討していたとき、「alvaDesc」でかなり高い性能の材料が発見できたなどの成果が出て、周りからも注目されまして、今に至るという感じです。

奥山さん：モデル化する際に使用する記述子の数は1,000～2,000

程度ですね。

中澤さん：スクリーニングを考えると、モデルができていればデータ数は100もあれば十分ですね。実験でデータを取るのがたいへんということもありますが、10個のデータでおおよその傾向がつかめることもあります。次の段階として、明らかに不要な候補を削除できるという点では有効です。

—ありがとうございます。「alvaDesc」で気に入っているところはどこでしょうか。

奥山さん：「alvaDesc」で一番すごいと思うのは、言葉が悪いかもしれませんが、第一原理計算と比べるとある意味粗雑な記述子じゃないですか。それで、ちゃんと予測までできてしまうのが驚異的だなと。確実に結果が出るところがいいですね。

中澤さん：わたしたちのところでは低分子も高分子も対象にしており、高分子はモノマー構造をもとにして扱っています。実際のデータを集めてくると、温度や条件などがまちまちだったりしますが、「alvaDesc」で計算すれば一律に同じ環境で得られたデータとして扱えるので安心して進められます。ある材料のこういう機能がほしいと思うときに、ある構造群を準備しておいて、ほしい機能に類似した物性値をもってきて代替させるようなことをすれば、似たようなものが得られる場合もあります。そういう使い方もできると思っています。

奥山さん：医薬分子を意識したような記述子が多いので、材料系に当てはめると、つくったモデルに対する解釈が難しい時はあります。ただ、以前は解釈も探索もできない時があったので、解釈が難しいながらも材料探索に有効だという意味では、足掛かりができることは大きいです。もはや、「alvaDesc」がないと研究が進まないですね。



押山さん

中澤さん：いや、実際に仕事ができなくなってしまう（笑）。ただ、どの記述子がいいのか、どういったものを伸ばせばいいのかはよくわかっていないんです。こちらで化合物をいろいろ準備して、それを「alvaDesc」で計算して、どれが高いか低いかを求めてそれを乗せてやるということであれば使えるなという感じです。そのあたりの理屈がある程度わかるようになると、計算に詳しくない人間でも使いやすく判定しやすくなると思います。

記述子分類わかりやすく、機械学習に向けた機能拡充を

—確かに、「alvaDesc」でメカニズム（機序）を知ることは厳しいですが、探索的なところではうまくご活用いただいているのですね。

中澤さん：報告する時は、何が効いているのかをはっきりさせることが大事ですから、「alvaDesc」のこの記述子ですというだけでは理解してもらいにくいかもしれません。こっちはそれでもいいのですが（笑）。それに、これだけいろいろな記述子があるのに活用しきれていないんじゃないか、使い方が十分わかっていないんじゃないかというような不安もあります。それで、ホームページなりで事例集のようなものを掲載していただけたら、ああこういう使い方もできるのかとか、参考にすることができると思います。

奥山さん：「alvaDesc」は記述子が多いため、その意味を調べるのが難しいと感じています。全部ウェブサイトに掲載するのもたいへんでしょうが、少なくともそれぞれの記述子の論文リファレンスとの対応がとれていると調べやすいと思います。記述子の名称でウェブ検索してもうまく探せません。あと、ソフトウェアとして詳しいマニュアルがあればうれしいですね。

中澤さん：計算科学があまり得意でない立場からすると、確かにもう少し親切だったらなと思いますね。記述子はおおまかに分類されていますが、何ができてどういう種類の記述子なのかピンときませんし、計算が専門ではないのでそれを調べる時間もあります。今後、われわれのような実験屋が利用するケースが増えると想定されているのであれば、そのあたり御一考願いたいです。

—こういう実験にはここが効いてくるはずだから、このブロックから選んでくださいみたいな、実験の現場でも理解しやすい分類がなされているとよいですね。ご意見ありがとうございました。事例集については、

御社の場合は材料系の用途が主なので、どれだけ参考になる例があるかわかりませんが、開発元に合わせて伝えておきます。ほかにも、ご要望等ありますか。

奥山さん：フィンガープリントの種類が増えるといいですね。いまはハッシュ化分子フィンガープリントですので、DENDRIC や Atoms-pair fingerprint などが使えるといいと思います。基本的に、いろいろなフィンガープリントを並べて評価してみたいので、たくさんできるほどうれしいですね。1 個だけのフィンガープリントでうまく予測できた経験がなく、いつも複数使います。総当たりで、いろいろなフィンガープリントの組み合わせを選びたいと思います。論文をみてもこういうやり方は多いです。

—わかりました。

奥山さん：ダミー変数が連続変数か、つまり 0 か 1 であらわされているのか、連続量であらわされているのかというカテゴリー分けもしてほしいですね。

機械学習を考えると、SMILES の文字列をそのままベクトル化する機能もほしいです。フィンガープリントを使ってしまうと、元の情報の一部が失われますから復号化できませんが、SMILES をベクトル化すると可逆的になり、数値から SMILES に戻せるメリットがあります。このプログラムを Python で書くことは実験屋さんには無理ですから、その機能が「alvaDesc」に付いていればいいなど。SMILES 文字列のベクトル化を用いて、新しい分子を機械学習により生成するということが、ここ数年の研究例ではみられます。そういうのに便利に使えるのではと考えています。

ステップ踏んで定着へ、「押しかけ MI」で普及活動

—実験の方々が非常によく活用されているのが印象的ですね。こうした土壌はどのように醸成してこられたのですか。

押山さん：もう 20 年くらい前の話ですが、計算科学を普及させようとした時に、実験の人たちからみると自分達の困りごとはどう役立つのかわかりにくいという声が多々ありました。それで、実験室や各開発室に「押しかけ女房」ならぬ「押しかけ計算科学」ということで、われわれが押しかけていって価値を理解してもらおうという活動から始めました。最初は門前払いされるようなこともあったのですが、だんだん事例をつ



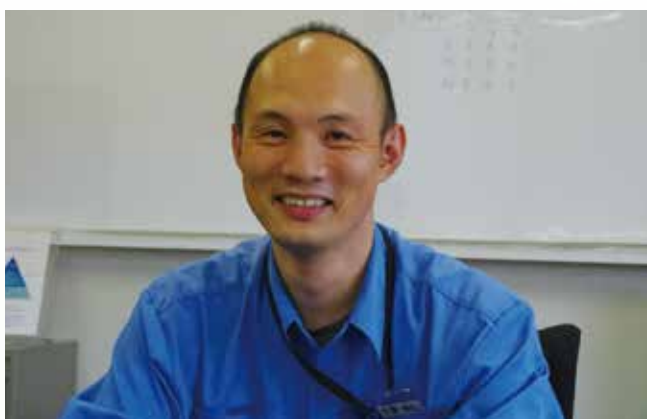
奥山さん



くっていったり、結果が出る仕組みづくりをしたりといった段階を踏んで、自立と定着を図るというステップで進めてきました。そのなかでは、大学の先生方から技術指導をいただいたり、外部の研究会に入って勉強したりもしましたし、計算と実験との人的交流や、専門知識を持つ人材採用なども行い、実験者が自ら計算を行う文化が育ってきたと思います。当社として、MIの取り組みは5～6年前から始めていて、現在はその社内普及を計算科学の場合と同じステップで進めています。まずは、実験室や各開発室に「押しかけMI」として出向き、共感してくれた人と事例をつかって、組織として定着させるためにどうするかを議論しています。

ー計算科学の時の取り組みがありましたので、MIの普及はスムーズだったのではないですか。

奥山さん：いやー、最初は提案してもピンと来ないと言われましたね(笑)。



中澤さん

押山さん：こんなのできないよとか、この処方ばちょっとおかしいんじゃないとか、そういう反応もありました。しかし、そのようなご意見は貴重です。それができそうにないならじゃあこうしてみようと考えますので、われわれとしても技術的に成長できる機会になりましたね。

中澤さん：データサイエンスの利用は、モデルさえきちんとしていれば、材料の性能もある程度読めますし、頭の中の演繹的思考が計算で具体化できるというメリットがあります。実験的方法だけだと、いろんな分析をして、たくさんデータを取って、結局何も出なかったということもありますので、時間を無駄にしないで済みます。

奥山さん：「alvaDesc」は簡単に計算ができるので、忙しい実験屋さんにはぴったりで、トライアル&エラーもやりやすいです。



セミナー・最新の情報は
ホームページをご覧ください



alvaDesc 製品ページ

無機系への適用に期待、クラウド対応も

ーそれでは、今後の予定についてお聞かせください。「alvaDesc」をこんなふうに使いたいとかございますでしょうか。

押山さん：いまの計算機環境は基本的に自前ですが、クラウドで利用できるものはそちらでの利用も進んできています。「alvaDesc」もクラウドにインストールできるなら考えたいですね。

ーサイトライセンスなら可能です。別途ご相談させてください。

奥山さん：もちろん、「alvaDesc」はこれからも使っていきますが、無機系への対応が課題になります。先ほども言いましたが、第一原理計算に比べて粗雑な感じでもいいので、無機系の記述子を用意していただくと助かりますね。

中澤さん：無機材料に対応した記述子は集めたいと思っていますので、同じ「alvaDesc」でやれるのならありがたいです。

奥山さん：いま、ポリマーもけっこう対象にしています。ただ、ポリマーをどのように定量化すればよいかという事は、まだあまり蓄積がないということもあるので、これから工夫して取り組んでいきたいと思っています。

中澤さん：われわれが取り扱っている材料は、大きく分けて樹脂と、可塑剤のような添加剤、フィラーなどですが、高分子は共重合体の組成の説明変数化が重要です。フィラーも説明変数化が必要ですが、残念ながら無機物のフィラーは現状「alvaDesc」では扱えず別の手段が必要になってしまいます。わたしは、「alvaDesc」に無機の記述子が導入されると、一元化することが出来るため、使い方が広がると思っています。

奥山さん：グラフィックプロセッサ(GPU)対応はどうなんでしょうか。ポリマーなどで探索候補の数が5万、10万になると、いくら計算が速いとはいえ時間がかかってしまいます。

ー「DRAGON」はシングルスレッドでしたが、「alvaDesc」はマルチスレッド対応になりましたので、そちらを活用していただくことができますが、今後、中分子への対応や力場による構造発生などの機能強化を考えたときに、GPUでの高速化が必要になるかもしれません。これは開発元と相談してみます。今日は長時間ありがとうございました。

(インタビュー実施：2021年9月)



Affinity Science

株式会社アフィニティサイエンス (英語名 Affinity Science Corporation)

住 所 〒141-0031 東京都品川区西五反田 1-11-1 アイオス五反田駅前

電話番号 03-6417-3695 FAX 番号 03-6417-3696

電子メール info@affinity-science.com

https://www.affinity-science.com/