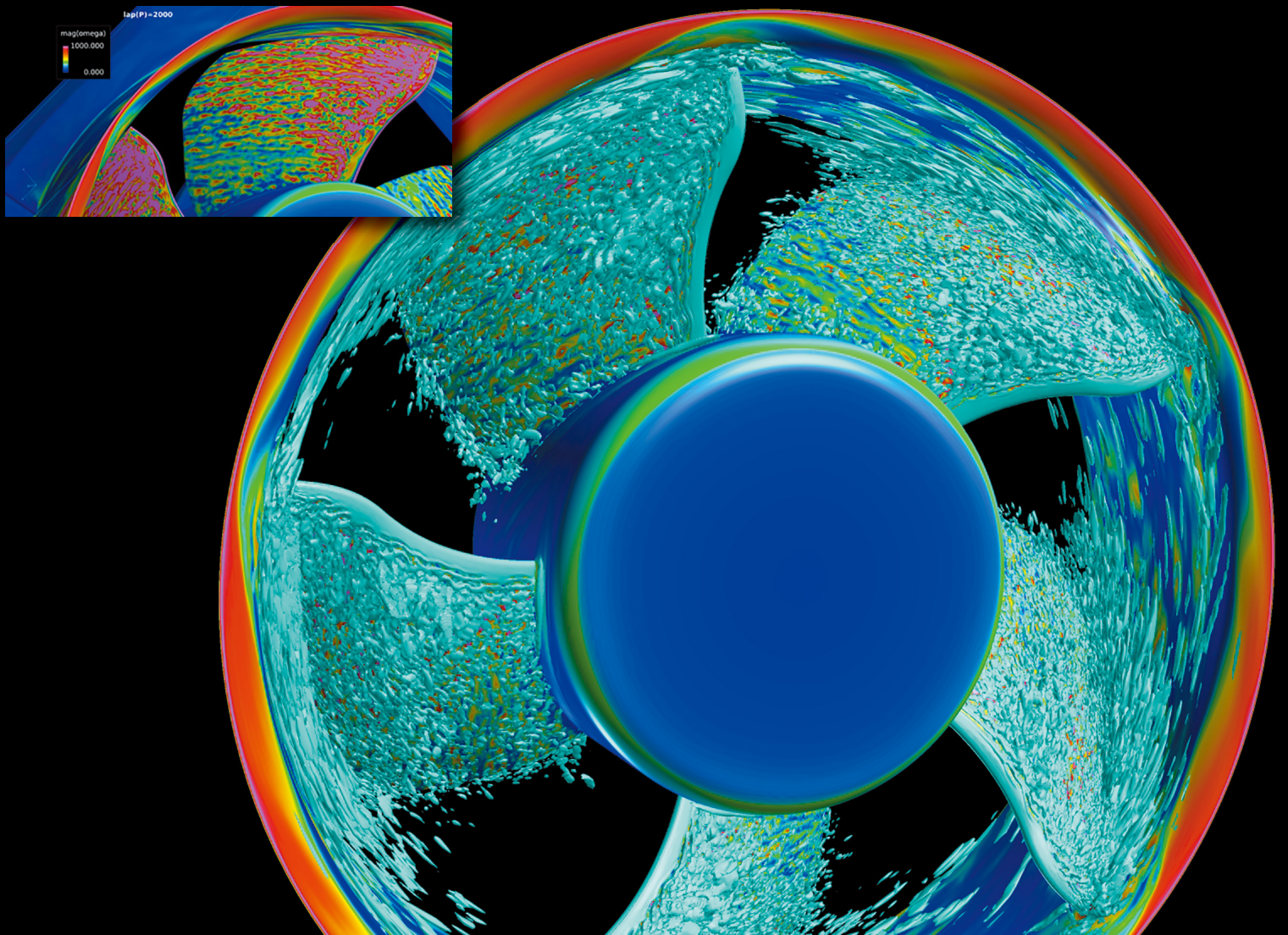


計 算 互 学 ナ ビ

計算工学ナビ・ニュースレター2019年秋号



特集 HPCと機械学習 産業界・アカデミアの先端事例

シミュレーションと機械学習を組み合わせたマテリアルズ・インフォマティクスにより革新的な材料開発を推進 …… 横浜ゴム株式会社 **小石 正隆**
 次世代半導体材料として注目されるシリコンカーバイドの開発にシミュレーションと機械学習を活用し開発期間の短縮化を実現
 …… 名古屋大学 **宇治原 徹**

特集 HPCと機械学習 ポスト「京」重点課題における機械学習の取り組み

ポスト「京」重点課題⑧ 近未来型ものづくりを先導する革新的設計・製造プロセスの開発
 ポスト「京」の時代のHPCとデータ科学 …… 東京大学生産技術研究所 **加藤 千幸**
 ポスト「京」重点課題⑥ 革新的クリーンエネルギーシステムの実用化
 磁場閉じ込め核融合プラズマの温度予測を機械学習で加速する …… 量子科学技術研究開発機構 **本多 充**

シミュレーションと機械学習を組み合わせた マテリアルズ・インフォマティクスにより 革新的な材料開発を推進



横浜ゴム株式会社 平塚製造所
研究先行開発本部 小石研究室

小石 正隆 理事・研究室長

横浜ゴムの小石正隆研究室長は、タイヤのゴム材料開発に役立つ情報を見出すため、進化計算と大規模動的粘弾性シミュレーションで得られたフィラー充填ゴムの微細構造(モルフォロジー)を定めるパラメータと等価物性値(計算結果)からなるデータセットを、自己組織化マップと機械学習を利用してデータマイニングするという「マテリアルズ・インフォマティクス(データサイエンスを活用した材料開発)」の推進に取り組んでいます。「成功へのカギは良質なデータと創造力、そして人間のひらめきも重要」と語る小石室長に、これまでの成果や課題などについてお話をいただきました。

材料設計にマテリアルズ・インフォマティクスを活用

最初に、タイヤ開発における材料設計の重要性や難しさについてお話しください。

弊社はタイヤの製造販売を行っています。タイヤ性能にとってゴム材料の役割は大きなウエイトを占めており、ゴム材料の研究開発も同時に行っています。タイヤのゴム材料には、転がり抵抗や摩耗などのタイヤ性能を改善するため、主成分の高分子にカーボンブラックやシリカなどのナノスケールのフィラーと呼ばれる充填剤が配合されています。ゴムの力学特性は、フィラーの分散状態などのモルフォロジー(微細構造)によって異なることが知られていて、フィラーをどのように生かすかが材料設計の重要な課題になっています。しかし、力学特性、つまり損失正接(損失係数。タイヤの転がり抵抗に寄与)や弾性率、強度などを同時に改善するためのモルフォロジーについての設計指針は、未だ明確ではありません。力学特性とモルフォロジーに関する有用な情報が得られれば、フィラー充填ゴムの設計を革新できると期待しています。そこで、私たちはマテリアルズ・インフォマティクスを活用した材料設計について研究を進めてきました。

マテリアルズ・インフォマティクスについてご説明ください。

マテリアルズ・インフォマティクスは、簡潔にいうとシミュレーションや機械学習を使って材料開発を高速化・高度化しようとする取り組みで、データサイエンスを活用した材料開発の新たなフレームワークとして期待されています。タイヤの構造設計では、以前からシミュレーションで取得したデータを使ったデータマイニングに自己組織化マップや機械学習を使う同様の取り組みをやってきました。同じ取り組みを材料設計でもやろうとしたわけです。私たちが期待する力学特性(物性値)を実現する

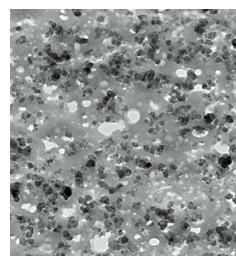
ために、シミュレーションでデータを取得して、機械学習でそのデータに隠れている原理・原則の理解や材料開発に役立つ情報を取得するという試みです。進化計算の枠組みのなかでゴム材料のモルフォロジーに関するパラメータを変更し等価物性値を算出するシミュレーションを数多く実施し、データを集め、等価物性値とモルフォロジーのパラメータとの相関関係を理解するために機械学習を活用したということです。タイヤ構造設計での実績があったので方法論は以前から分かっていたのですが、それを実現するためにはいくつかの障害がありました。つまり、モルフォロジーを定めるパラメータとは何なのか、また、そのパラメータに基づいてシミュレーションモデルをどのように構築すればいいのか、さらに、数多くの大規模シミュレーションを如何に効率的に実行するのかといった技術的な課題です。そうした課題を解決して、ようやく3年ほど前から本格的にマテリアルズ・インフォマティクスに取り組めるようになりました。構造設計のとき、最初はデータを自己組織化マップで可視化してその可視化情報を人が主観的に分析するやり方でしたが、それだと慣れていない人には分かるけれど、慣れていない人にはよく分からないわけです。それなら誰にでも分かるルール、つまり「ある特性値を高くするには、このパラメータをいくつかからいくつかの間にする」といった設計指針を機械学習で見つけ出せばいいのではないかと、そう考えて機械学習を使用したデータマイニングを行うことにしました。

人の気づきやひらめきに期待

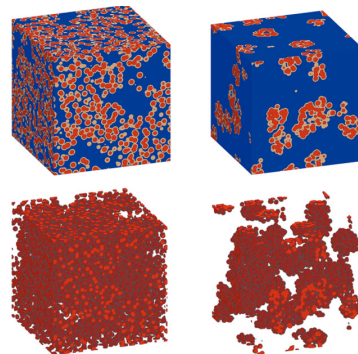
マテリアルズ・インフォマティクスによるフィラー充填ゴム設計についての研究を進める上で、最も苦労されたのはどんなところでしたか。

たくさんありましたが、先ほどもお話ししたように私たちの取り組みのなかで大きな

課題が2つありました。1つ目は、モルフォロジーのパラメータ化とマルチスケールランダムモデルというモデリング手法によるシミュレーションモデルの構築、もう1つは、1億要素以上の大規模シミュレーションを、進化計算のなかで数千回実行しなければならないため、これをいかに効率的に行うかということでした。シミュレーションの役割はいろいろあると思いますが、機械学習に活用するという立場からすると、その大きな役割は仮想的な条件でのデータ生成です。条件を変えてたくさん計算すれば、製造設備など実世界での制約を度外視した多



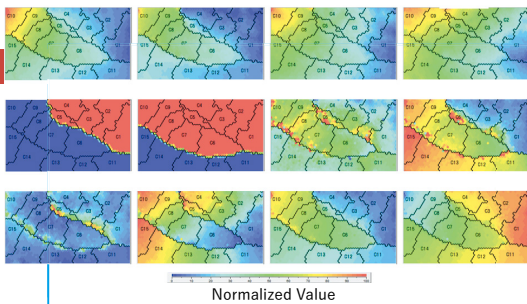
▲カーボンブラック充填ゴムの電子顕微鏡写真



領域サイズ: 500x500x500nm,
要素数: 1億3千4百万(512x512x512)

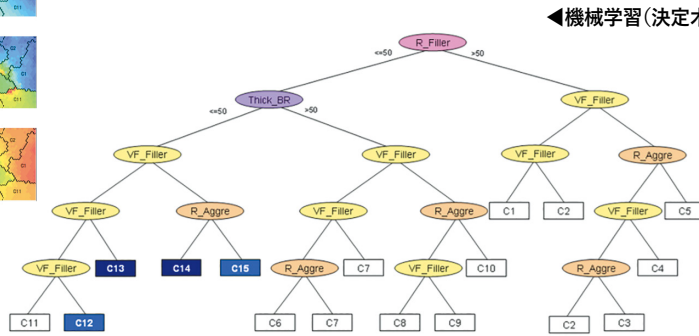
▲フィラーのモルフォロジーの異なるシミュレーションモデル例

左は分散が良いとされるモルフォロジー、右は凝集塊が多く分散が悪いとされるモルフォロジー(フィラーの半径と体積分率は同一)。上段は全体を、下段はフィラーのみ表示している。



▲目的関数と設計変数を正規化した値で可視化した自己組織化マップ

自己組織化マップによって2次元平面上に写像された全データを15のクラスターに分類し、さらに目的関数と設計変数それぞれの数値でコンター表示している。目的関数は上段左から等価弾性率 (Modulus)、等価損失正接 (Loss Tangent)、平均応力 (AVE_Stress)、応力の標準偏差 (STD_Stress) を表す。4つの目的関数がバランスした最も好ましい設計領域として C13 と C14、これに準ずる設計領域として C12 と C15 を選択した。



- ▶ フィラー半径で分岐し、目指すラベル(C12-15)はフィラー半径が小さい方に存在
⇒**フィラー粒径を小さく**
- ▶ バウンドラバーの厚さで分岐し、目指すラベルはバウンドラバーが薄い方に存在
⇒**バウンドラバーを薄く**
- ▶ さらに、フィラーの体積分率と存在領域半径(分散状態)で細分化

自己組織化マップで得られた15のクラスター (C1~C15) を結論として、8つの設計変数を説明変数とした分類ルール (木構造) を機械学習で見出した。4つの好ましい設計領域 (C12~C15) の位置から、4つの目的関数がバランスしたフィラー充填ゴムを設計するには、フィラー半径を小さくし、バウンドラバーを薄くすることが最も重要であり、さらにフィラーの存在領域半径とフィラーの体積分率をコントロールすべきであることが確認できた。

くのデータを生成できます。そのデータを上手に使える、そこから新材料開発に役立つ情報が得られるわけです。このときに大事なのが、設定した設計空間でのデータのサンプリング方法 (実験計画法) です。直交実験やラテンハイパーキューブによるサンプリングが常套手段ですが、最適なところはなかなか出てこなくて、だいたい外すものなんです。そこで、最適解を含むその周辺のデータを効率的に取得するために進化計算をデータのサンプリング手法として使うことにしました。タイヤやゴム材料の最適化は目的関数が複数あるため多目的最適化問題になりますが、進化計算で多目的最適化を計算すると、パレート解という最適解に近いデータを見つけ出すことができます。

通常のサンプリングでは、最適解を外すことが多いのですか。

経験的に最適解は設計空間の端にあることが多く、うまく端を捉えるサンプリング方法がないため、サンプリング数を増やしても最適解を含むデータを得るのは難しいですね。また、そもそも設定した問題に望む答えが存在するののかという問いを繰り返すこととなります。はじめに目的やパラメータや目的関数を決めて設計問題 (多目的最適化問題) を設定するのですが、そこに必ず望む解があるとは限りません。私たちは、機械学習から得られた知見を活用して、もう一度問題設定をやり直すということをよくやります。そのためのヒントが、機械学習が出した答えのなかに含まれていることもあるのです。

つまり、問題設定をするのは人であり、人がどのように関わるかが重要ということですか。

マテリアルズ・インフォマティクスといっても、コンピュータが自動的にすべて材料設計をしてくれるわけではありません。私は、シミュレーションや機械学習は、人が材料設計するための支援技術と考えています。一方、人は常に合理的に判断し、間

違うことがないかということ、全くそんなことはなく、同じ人でも状況によって判断が変わったり、間違ったりするものです。だからこそ、その原因であるバイアスなどの存在を前提として、シミュレーションや機械学習をツールとして活用しながら、人を中心に据えた設計開発を進めていきたいと考えています。私は、よく「人の気づきやひらめきに期待している」という話をします。多くのデータをもとに機械学習をやれば、改良に役立つ情報が出てきます。私たちが目指していることが、今ある材料の改良でよいなら、その情報で問題なく進めていけるかもしれません。でも、今の材料とは違う、新しいものを発見したいと思ったら、その答えはおそらく今の材料の延長線上にはありません。すでにある領域のデータだけを眺めていても新しい発見は生まれません。では、どうすればいいのかというと、機械学習から得られた情報を理解した上で、さらにそこからの発想のジャンプ、つまり気づきやひらめきによって、今までの延長線上にないデータを生み出す設計問題へと変更する、そういうアプローチを続けていくことが必要です。その意味で、私は人の可能性に期待していますし、「人間中心で進めていこう」と話すのです。目の前の高い壁を打ち破るのはシミュレーションや機械学習ではなく人です。得られた情報から、壁を破るための新たな問題設定ができれば、やがて新たな発見や不連続なイノベーションにつながると考えています。また、失敗データも壁を破るためには重要です。

シミュレーションと機械学習はいいコンビ

これから機械学習に取り組んでいこうと考えている人たちも多いと思います。そんな人たちにアドバイスをお願いします。

間違いなくいえるのは、新たな発見を期待する場合、シミュレーションと機械学習

はずごくいいコンビだということです。今年の2月に発表された経団連の提言書 (AI活用戦略) でもシミュレーションと機械学習の活用を勧めていますので、それは確かです。設計開発のためにシミュレーションをやっている方は、ぜひ機械学習を活用してほしいと思います。ただ、よく聞かれるのは、「シミュレーションと機械学習をどう組み合わせたらいいのかよくわからない」という声です。我々がどうやってきたかという、進化計算の枠組みでパラメータを変更してシミュレーションを行い、パラメータと計算結果が対になったデータを取得し、機械学習を使ってその相関関係や因果関係を理解するというやり方です。目的関数が複数あるため、データを自己組織化マップでクラスタリングし、そのクラスターラベルに関する知見を決定木で導いています。深層学習ではなく決定木を使うのは機械学習の判断理由を知りたいからです。もちろん機械学習には他の使い方がありますが、この活用方法は設計開発にとって非常に有効だと考えています。

マテリアルズ・インフォマティクスによるフィラー充填ゴム設計に関するご研究は、今後、どのように進めていかれるのですか。

これまでの実証実験で、1億3400万要素からなる大規模動的粘弾性シミュレーションを進化計算に基づいて6,000回以上実施しました。自己組織化マップと機械学習を利用して、6,000件のデータセットをデータマイニングした結果、フィラー充填ゴムのモルフォロジーを定める重要なパラメータに関する情報が得られました。今後は得られた情報の検証を行い、よりよいゴム材料をつくるための次のステップへ進めています。このほかにも、例えば転移学習によるシミュレーションデータと実験データの融合など、やるべきことはたくさんあります。やるべき研究の1つの段階には到達できましたが、まだ終わってはいません。

次世代半導体材料として注目されるシリコンカーバイドの開発に シミュレーションと機械学習を活用し 開発期間の短縮化を実現



名古屋大学 未来材料・システム研究所
未来エレクトロニクス集積研究センター
宇治原 徹 教授

パワーデバイス用の省エネ新材料として期待されるシリコンカーバイド(SiC)の高品質化をめざして、名古屋大学の宇治原 徹教授らは、溶液成長法と呼ばれる結晶成長手法による材料プロセス開発に取り組んでいます。半導体材料として広く利用されているシリコンも大口径高品質結晶成長技術の確立までに約40年かかったように、新材料のプロセス開発には長い年月が必要とされています。宇治原教授はシミュレーションと機械学習をフル活用し、この開発期間を通常の10分の1以下にまで短縮することを可能にしました。機械学習を研究開発にどのように役立っているのかなど、これまでの成果についてお話しいただきました。

パワーデバイス向け新素材として 期待されるSiC

先生が研究開発に取り組んでおられるSiCとはどのようなものですか。

SiCがパワーデバイス（電力変換器などに使用される半導体素子）用の新素材として優れた特性を持つことが知られるようになったのは20年以上前のことです。SiCの大きな特長は、変換時のエネルギー損失が少ないことです。現在広く使用されているシリコン製のデバイスをSiCにすると、損失が約3分の1になるといわれています。SiCは革新的な省エネルギー材料として、すでに一部の電化製品をはじめ、太陽光発電機器、鉄道などで実装されており、次の新幹線にもSiCが使われるだろうといわれ、さらに今後の電気自動車の増加などを考えると需要はますます拡大することから、SiCへの期待は高まっています。ただし、課題は結晶の品質です。品質が悪ければ、当然ながら高い性能は得られません。現在、SiCは「昇華法」という手法によってつくられています。例えばシリコンの場合は、るつぼのなかで溶かして、それを引き上げるというやり方で結晶成長させますが、SiCの場合は溶けずにいきなり昇華します。ドライアイスのように液体状態がないわけです。原料の粉末を2,000℃以上に加熱すると、すべてガスになり、上部で少し温度を下げておくと、そこで再結晶化します。そうやって結晶成長させるわけです。しかし、この方法では高品質の結晶をつくり出すことが難しいのが実状です。転位がたくさん含まれてしまうのです。転位とは、分かりやすくいうと“原子のずれ”です。転位があるとリーク電流が漏れ出し、性能

がどんどん低下します。この転位の原因は温度差にあります。温度差によって熱膨張に差が生じてしまい、そのつじつま合わせで原子がずれてしまうわけです。「昇華法」では、昇華したガスを再結晶化させるために、必ず温度差が必要ですから、どうしても転位が多くなります。

そこで先生は、別の手法によって高品質のSiCの結晶成長を実現させる研究を進めてこられたわけですね。

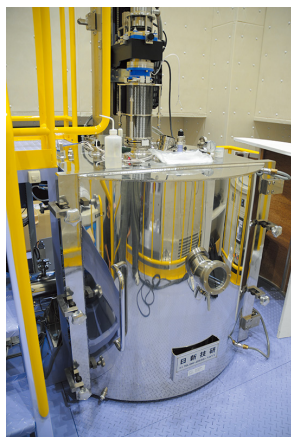
それが「溶液成長法」です。まずカーボンのるつぼのなかでシリコンを溶かします。するとシリコンの液体のなかでカーボンが溶け込み、過飽和状態となってごく小さな温度分布のなかで結晶を成長させることができます。その結果、転位密度が低い高品質の結晶が得られるのです。この方法は、現在、ほとんど誰もチャレンジしてなくて、私たちは間違いなく世界でいちばん高品質のSiC結晶をつくる技術を持っているといえます。しかし、問題はその先です。いくら高品質の結晶成長技術を確立したとって

も、それは10mmとか15mmという、まだすごく小さな結晶でしかありません。実用化に向けて大型結晶を実現させるためのプロセス開発にはまだ時間が必要です。ちなみに、シリコンの場合、高品質結晶成長技術が確立したのが1958年で、大口径化が完成したのが1995年ごろです。およそ40年かかっています。しかし、SiCの開発に40年もかけるわけにはいきません。そこで私たちは、シミュレーションと機械学習をフル活用して、研究開発期間の短縮をめざすことにしたのです。

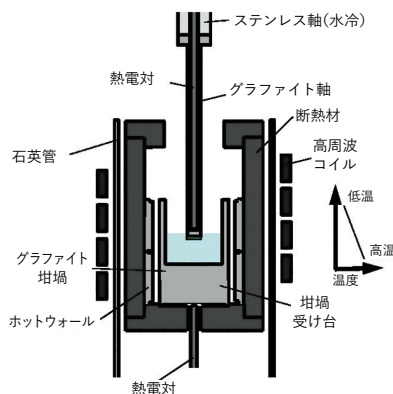
機械学習で結晶成長を 精度よく高速で予測

どのようにシミュレーションと機械学習を活用したのですか。

研究成果を上げるためには、液体中の温度分布や組成分布、流れの分布などを思い通りに制御することが求められます。しかし、1,800~1,900℃という高温のるつぼのなかでどうなっているのかは見ることはできません。そこでシミュレーションを利用しました。し



▲宇治原教授らが開発した結晶成長装置



結晶成長装置 内部の仕組み

カーボン製のるつぼのなかでシリコンを溶かし、カーボンスティックの先端からSiCの結晶を成長させる。このとき、るつぼやスティックを回転させたり、液体の量などをさまざまな条件を変化させて、るつぼ内の温度分布や流れ分布などを制御する。

しかし、制御するパラメータがあまりにも多過ぎて、最適条件はそう簡単には見つかりません。例えば、るつぼの温度を何度にするか、どのように加熱すればよいか、るつぼやそこに差し込むカーボンスティックをどのように動かしたり回転させればよいか、さらにはるつぼの厚みや形状をどうするか、るつぼを覆う断熱材の厚みはどうするかなど、パラメータは挙げればきりがなく、その組み合わせは無数にあります。シミュレーションがいくら速いといっても、やはり限界があります。そこで機械学習を用いて、結晶成長装置内部の状態を瞬時に予測し、可視化するシステムを開発しました。このシステムによる予測結果をシミュレーションで計算した結果と比較するとほとんど違いがなく、非常に精度よく高速で予測できていました。さらに、予測から最適条件を効率的に見つけ出すことも可能になろうとしています。つまり、今までは「パラメータがたくさんあり過ぎて計算できない」ということができたようになり、さらに、ニューラルネットワークを使って簡単に最適値を求められるようになりました。実際に今、私たちは構築したニューラルネットワークによって目的関数をしっかり定義した上で、たくさんのパラメータを最適化することにも取り組んでいます。

その結果、実験のやり方も変わってきました。今までは少ないシミュレーション結果から、「こんなものかな。これでやってみるか」と実験条件を決めていましたが、今は最適化ツールを使うと、パーツと最適化が始まって、すぐにビューアーに映し出されて、その結果を見ながらみんなでディスカッションして「よし、これでいこう」と、ある程度確信をもって実験に取り組めるようになりました。また、機械学習を使うと、ひとつのこと

を実現するにもその方法が何通りかあることを教えてくれるので、実験を行うときの条件の決め方もすっかり変わりました。1回の実験を実施するには2日間ほどかかりますから、この差は大きいです。さらに、実験装置の大きさや形状まで変えらるとなるとどんどん条件が変わってしまうため、今まではなかなか提案しにくかったのですが、今は「るつぼのサイズを変えるとどうなるか」、「断熱材の厚さを薄くするとどうなるか」までサジェスチョンしてくれるので、そうした変更もしやすくなりました。こうして、シミュレーションと機械学習、そして実験を効果的に組み合わせることで研究開発を進めるようになってから、2年ほどが経ちました。かつては10mmの高品質結晶しかできませんでした。2年ほどの間に85mmまで大きくすることが可能になりました。今では150mmの大型結晶の実現も、そう遠くないところまで研究開発は進んでいます。シリコンでおよそ40年かかった開発期間を10分の1ほどに短縮することも夢ではありません。

実験によって目的関数を しっかり決めることが課題

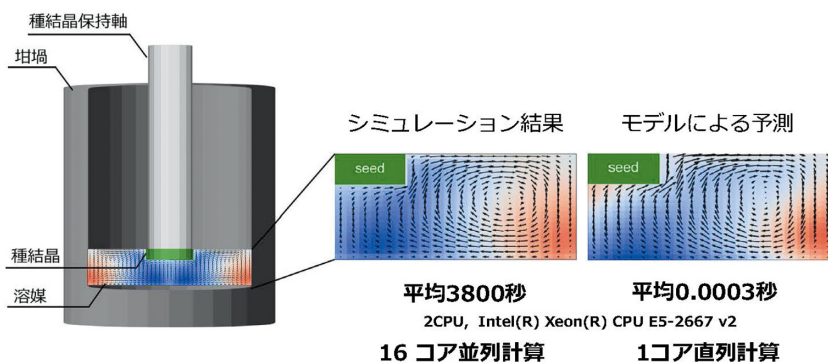
機械学習を活用した研究開発で、課題になっていることはありますか。

結晶成長の基礎研究から実用化研究までの進化のスピードを驚くほど速くすることが可能になったわけですが、今後の課題と思っているのは、目的関数がまだしっかり決められていないということです。基礎研究レベルで解明しなければいけないことがまだ残っています。シミュレーションと機械学習と最適化という方法は確立できましたが、結局、最適化しようと思ったら、目的関数というか最適化条件をきちんと決めなければいけない。

そのためには、まだ多くの実験が必要なのだろうと考えています。具体的にいうと、結晶成長速度が足りないのです。実用化のためには成長速度を高めるにはどうすればよいかを明らかにしなければならず、そのためには、やはり実験が重要です。

機械学習については、今後どのように進めたいかをお考えですか。

多くの人から聞かれるのは、「最適化に使ったモデルは、本当に合っているのですか」ということです。「間違っていたら、何の意味もないですよ」と。確かにその通りですが、もしシミュレーションが間違っていたら、その原因は、支配方程式そのものが間違っているか、扱っている物性値の正しい値がわからないから間違っているという2つの可能性があると思っています。私たちは、自分たちが開発した機械学習モデルを使って、実際の実験結果と比較することで、物性値が正しい値かどうかを検証する取り組みも進めています。この方法を使えば、機械学習によって今まで誤っていたかもしれない物性値を正すこともできます。さらに、シミュレーションを再現するモデルを実験と比較することで、シミュレーションモデルの間違ひも正すことができるようになりました。つまり、実験をやればやるほどデータが蓄積されるといったレベルではなく、実験を重ねることによって正しいモデル、正しい物性値が明らかになる、そういう仕組みをつくったわけです。こうした手法は、今後はSiCだけでなく、現在名古屋大学を中心に研究開発が進んでいる窒化ガリウム (GaN) のプロジェクトなどにも応用することができます。さらにその他の新素材の結晶成長プロセスや熱処理プロセスにも応用することも可能であると考えています。



▲機械学習を用いて、容器内部の状態を高速に予測

SiC結晶成長時のるつぼ内の温度分布(色)や流れ分布(矢印)を機械学習を活用して予測。その予測結果(右)とシミュレーション結果(左)を比較すると、ほぼ一致し、予測の精度が極めて高いことが示された。



▲結晶成長装置の内部を瞬時に予測してアニメーション化し、プロジェクションマッピング技術により装置に投影。臨場感のある映像により、オペレータは内部の状態を把握しながらパラメータを制御することができる。



ポスト「京」重点課題⑧
「近未来型ものづくりを先導する革新的設計・製造プロセスの開発」

ポスト「京」の時代のHPCとデータ科学

加藤 千幸 センター長・教授 ポスト「京」重点課題⑧ 課題責任者
東京大学生産技術研究所 革新的シミュレーション研究センター

はじめに

2012年9月28日に共用を開始した「京」の運用が8月16日を以って終了しました。「京」は我が国のHPCI (High Performance Computing Infrastructure) の中核スパコンとして、この7年近くの間にはわたりHPCを強力に牽引し、数々の革新的な成果を上げてきました。一方、「京」に代わる次期スパコン(フラッグシップマシン)として、アプリケーションで最大「京」の100倍の性能を発揮することを目標にして、2014年に開発が着手されたポスト「京」(スパコン「富岳」、以下「富岳」)は今から1年半後にあたる2021年4月の運用開始を目指して、急ピッチで開発が進められています。重点課題⑧では、理化学研究所計算科学研究センター(R-CCS)、および、富士通株式会社と強力に連携することにより、「富岳」において高性能を発揮できる、ものづくり分野のキラーアプリケーションの開発を進めています。この記事では、アプリケーションの開発状況をご紹介します。1年半後に迫ってきた「富岳」の時代におけるHPCについて言及するとともに、HPCとデータ科学との融合に関しても展望します。

コデザインによるハードウェアとアプリケーションの協調開発

「富岳」のプロジェクトの大きな特長はハードウェアとアプリケーションの協調開発です。まず、各重点課題で、ターゲットアプリケーションとよばれる合計9本のアプリケーションを選定し、これらの全てのアプリケーションでバランス良く性能を発揮する

ように「富岳」は設計されています。重点課題⑧では、有限要素法による非圧縮性流れ解析ソルバであるFrontFlow/blue (FFB) がターゲットアプリとして選定されており、それ以外の3本の主要アプリケーションの開発者も参加して、アプリケーションの開発を進めています。現在のアプリケーションの開発状況を表1に示します。「京」のノードと「富岳」のノードの理論ピーク性能はそれぞれ128 GFLOPS、2.7 TFLOPS以上であり、ノードの理論ピーク性能比は約20倍ですが、これらの4本のアプリケーションでは10倍以上の高速化が実現できる見込みが得られており、一部のアプリケーションでは、ノードの理論ピーク性能比以上の性能向上が図られる見通しです。たとえば、ターゲットアプリであるFFBに関してはさらなるアルゴリズムの改良により、この5倍程度の高速化が実現できる見通しが得られています。1年半後に迫った「富岳」の時代には、総じて現在よりも数10倍以上高速にアプリケーションが実行できる環境が整い、HPCの実用化が飛躍的に進展することが期待されています。なお、表1において、FFB以外のアプリケーションの「富岳」のノード性能はシミュレーター^[1]による推定値であることを付記します。

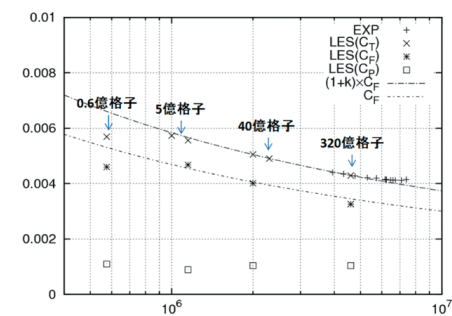
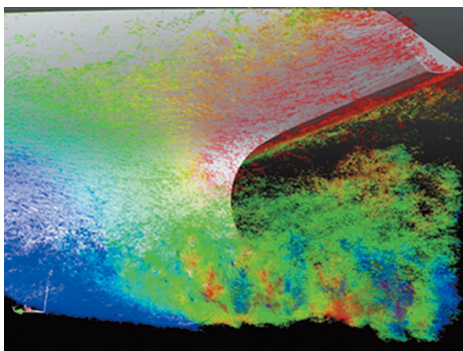
アプリケーションの高速化によるHPCの実用化の加速

「富岳」の時代にはこれらの高速化されたアプリケーションによりHPCの産業応用が加速的に進むことが期待されています。たとえば、図1は船の粘性抵抗を大規模な流体解析により予測したものです。これ

以外にも合計5隻の船の粘性抵抗を数値解析により予測し、試験結果と比較したところ、いずれの船に関しても数値解析により粘性抵抗を高精度に予測できることが確認され、将来的には曳船水槽試験を数値解析により完全に代替できることを示すことができました。ところが、この計算には「京」の2万ノードを用いました。そのため、計算ノード数が多過ぎ、直ぐには実用化するには至りませんでした。「富岳」の時代にアプリケーションの性能が10倍から100倍程度に高速化されれば、必要な計算ノード数を10分の1から100分の1に削減することができ、このような大規模計算の実用化が進展することが期待されているわけです。

HPCとデータ科学の融合

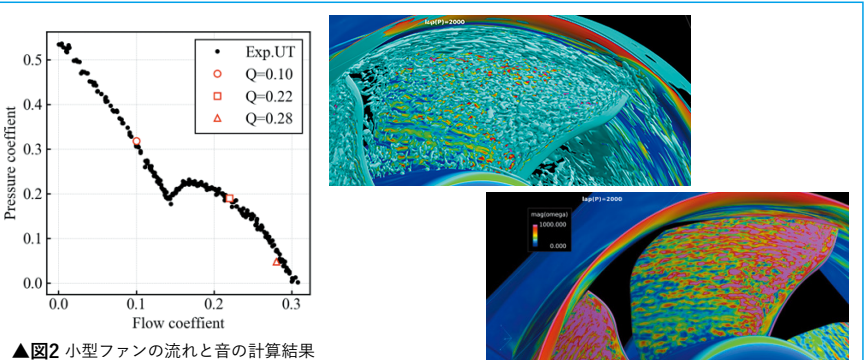
アプリケーションの性能が数10倍以上高速になる「富岳」の時代には、HPCにより大量のデータが生成されます。しかも、これらのデータの多くは、経験則を含まない第一原理的な手法により生成されるため、従来よりもはるかに高精度なデータです。このような大量かつ高精度なデータを利用することにより、データ科学が急速に進展することが期待されています。たとえば、ものづくり分野におけるHPCとデータ科学の直接的な応用を考えると、設計変数、性能を決める主要な物理現象、性能との間の相関関係を機械学習などのデータ科学の手法を用いて分析することにより、設計変数と性能との間の直接的な相関関係を見出すことができる可能性があります。このような場合には、物理現象のシミュレーションを介して製品の性能を予測していた設計プロセスを、設計変数から直接高精度に性能を推定できるようになり、設計の精度とスピードが大幅に向上することが期待されます。たとえば、図2は性能と騒音を目的関数として小型のファンの動翼形状を遺伝的アルゴリズムを用いて最適化している研究ですが、この最適設計の過程で得られる大量のデータを分析し、設計変数と性能との間の直接的な因果関係が、今よりも高度に理解できれば設計自体が高度化することが期待されます。また、大規模な計算をする代わりに、簡易なモデルで現象を予測する手法であるモデルリダクションにおいても、HPCにより生成されるデータとデータ科学の分析手法を活用することにより、従来よりも高精度な簡易モデルを構築することが可能になります。筆者らのグ



▲図1 320億の計算格子を用いた船のまわりの流れの高精度計算

アプリケーション	機能・特長	「京」のノード性能	「富岳」のノード性能	性能向上比
FFB	FEMによる非圧縮性/圧縮性LES解析・超大規模な高精度解析	4.0 GFLOPS	85.2 GFLOPS	21.3倍
CUBE	FVMによる非圧縮性/圧縮流体解析・直交格子による完全自動格子生成とオイラー系連成解析	18.3 GFLOPS	234.6 GFLOPS	12.8倍
FFVHC-ACE	FVMによる超音速流体解析・レイヤー格子と直交格子のハイブリッド計算・高精度な壁面モデル	6.2 GFLOPS	130.0 GFLOPS	21.0倍
FrontISTR	非線形構造解析・高収束性前処理型マトリックスソルバ	9.2 GFLOPS	106.4 GFLOPS	11.6倍

▲表1 ポスト「京」重点課題⑧のアプリケーションの開発状況



▲図2 小型ファンの流れと音の計算結果

ループでは、前述の船の計算に機械学習を応用することを考えています。アプリケーションの高速化により数値曳航水槽の実用化することは可能だと考えていますが、粘性抵抗を試験と同じ程度の精度で、かつ、短時間に予測することができるになれば、船型設計自体を飛躍的に高度化できます。粘性抵抗を決めているのは、船体表面近くに無数存在している1 mm以下の微細な渦の運動ですが、HPCを利用すればこ

のような微細な渦の運動に関する正確なデータを取得することができます。データ科学的な手法により渦の運動を分析し、その理解が進めば、従来の限界を超えた乱流のモデル化ができる可能性があります。現在、そのための予備的な検討を進めています。

■ おわりに
1年半後に運用が開始される見込みである「富岳」

の時代には、今の数10倍以上高速な計算が可能ならアプリケーションが多くなり、HPCの産業応用が加速度的に進むことが期待されています。また、HPCにより生成される大量でかつ高精度なデータが利用できるようになり、データ科学も飛躍的に進展することが期待されます。今後、HPCとデータ科学との結びつきはますます強まることが予測され、さまざまな分野においてその時代に向けた準備研究が進んでいます。



ポスト「京」重点課題⑥
「革新的グリーンエネルギーシステムの実用化」

磁場閉じ込め核融合プラズマの温度予測を機械学習で加速する

本多 充 上席研究員 量子科学技術研究開発機構 核融合エネルギー部門 那珂核融合研究所 先進プラズマ研究部 先進プラズマモデリンググループ

■ 高度な輸送モデルでプラズマの温度分布を予測

磁場閉じ込めプラズマの研究において、プラズマ温度分布を正確に予測することは極めて重要です。定常状態での温度は核融合反応や加熱装置による加熱と、熱をプラズマ外へ運び出す機構である輸送が釣り合うことで決定されます。輸送の主な駆動源はプラズマ中の電磁場の揺動に起因する乱流ですが、乱流輸送を正確に記述する事が可能な大規模シミュレーションは計算コストが大きいため、ルーチ的に温度予測する用途には向きません。そのため乱流輸送の素過程をモデル化した輸送モデルを、1次元拡散型輸送方程式を解く輸送コードに組み込んで温度を予測することが一般的です。乱流輸送は温度や密度などの勾配に強く依存しており、臨界勾配値を超えると急速に輸送流束が増大する非線形の性質を持っているため、工夫無しには計算が上手く収束せず振動解や発散解となってしまいがちでした。

この問題を本質的に解決すべく、輸送方程式を一般的な空間離散化手法で数値的に解くのではなく、輸送流束と加熱による流束が一致する温度と温度勾配の組み合わせを大域的最適化手法によって直接探索する、全く新しい輸送コードを開発しました。両流束が一致するときの温度や温度勾配は、定常状態における輸送方程式の解と対応しています。温度勾配値そのものを従属変数として取り扱うため、温度勾配値を評価するための数値微分も不要であり、また空間離散化に起因する数値的不安定も回避することが可能となります。数値的不安定を引き起こすことなく、高度な輸送モデルを使ったプラズマの温度分

布予測が実現可能となりました。実装では、大域的最適化手法の中でも多次元問題に強く安定に解を求められることが知られている遺伝的アルゴリズムを用いていますが、一方で膨大な回数の目的関数の評価が必要となる欠点もあります。現在の高度な輸送モデルは1回の輸送流束の評価に時間が掛かるため、輸送モデルそのものが並列化されています。計算を加速するために、遺伝的アルゴリズムにおいて個体ごとの並列化を行い、各個体が並列化された輸送モデルを呼び出すという、MPMDの枠組みで数値実装を行いました。典型的には2,000並列超で実行されます。最終的な解を得るまでに1,000万から1億回程度の輸送流束の評価が必要となることが多く、数時間から数十時間の計算時間が掛かってしまうことが問題でした。

■ 機械学習を用いた輸送モデル模擬による計算高速化

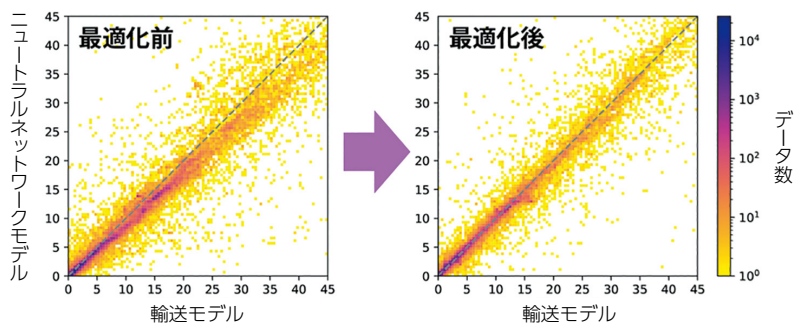
遺伝的アルゴリズムは局所最適を避けるために最適から遠い解も探索するため計算コストを増大させますが、このことは輸送モデルに対する幅広い入力パラメータ空間において大規模なパラメータサーベイを行っていることと等価となります。すなわち膨大な数の入力と出力の組み合わせを生成しているため、このデータを元に輸送モデルの振る舞いを模擬する代理モデルを構築し、それを元の輸送モデルと代替することで計算を加速させることとしました。代理モデルの構築には全結合ニューラルネットワークである多層パーセプトロンを用いました。入力・出力層のユニット数は輸送モデルに対する入力・出力と概ね等しくなるため比較的一意に決定されますが、隠れ層の数やそのユニット数、汎化性能を向上させるドロップアウト

率などのハイパーパラメータは経験的に決定する必要があります。学習に用いたデータはおよそ97.6万であり、検証・テストに用いたデータはおよそ10.8万もあります。学習そのものはGPGPUを用いて数分程度で終了します。遺伝的アルゴリズムは真の解からほど遠い多くの外れ値も生成しますが、実際の計算に対して意味を持つのは真の解に近い場所における代理モデルの再現性ですので、2標準偏差内に入る出力データに対してイオン熱流束の決定係数を算出したところ、0.897となり良好な結果を得ました。

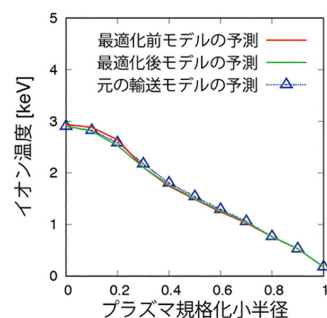
■ ハイパーパラメータの最適化

経験知に依らないハイパーパラメータの決定を行うため、ベイズ最適化手法の一つである、Sequential Model-Based Optimizationによるハイパーパラメータの最適化を試みました。目的関数を近似する確率モデルに、Tree-structured Parzen Estimatorを用いました。この手法の優れたところは、数値コストの大きい目的関数を直接評価するのではなく、計算の軽い目的関数の近似モデルを構築し、最適解探索に用いることで高速化を図っていることです。100回の試行で8時間ほど掛かりますが、最適化によって得られた隠れ層数、各層のユニット数、ドロップアウト率を元に学習した代理モデルは、イオン熱流束に対して0.943の決定係数を与えます。最適化後の代理モデルは最適化前よりも高い精度で元の輸送モデルが与えるイオン熱流束を予測できていることが、図1からも明らかに分かります。

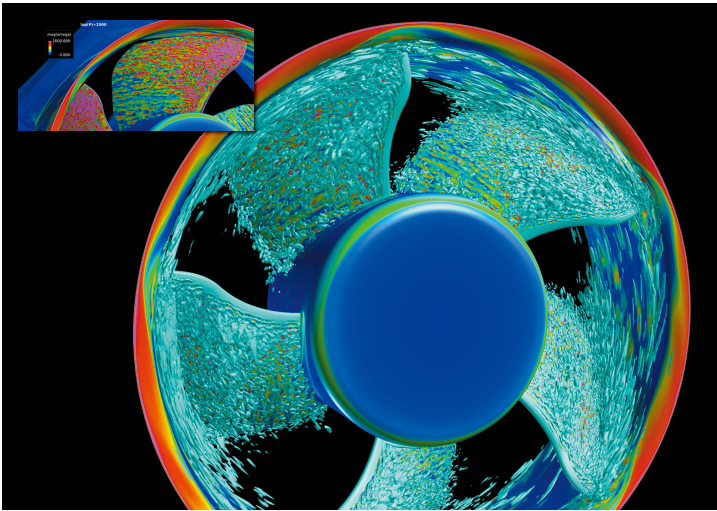
推論時に用いられるニューラルネットワークと同じ構造を持つサブルーチンをFortranで作成し輸送コードに実装することで高速な代理モデル計算が可能となりました。これらの代理モデルを用いて輸送コードで定常イオン温度分布を予測したものが図2で、元の輸送モデルを用いた場合の結果も載せています。両モデルとも良好な再現性を示していますが、最適化後の方がより再現性が高いことが分かります。一旦代理モデルを構築してしまえば、輸送コードによる温度分布予測に掛かる時間は数分で済むため、劇的な計算の加速を実現することが出来ました。機械学習手法を用いた高速なプラズマ予測手法の開発により、核融合炉の性能予測研究が加速することが大いに期待されます。



▲ 図1 最適化前(左)と最適化後(右)のニューラルネットワーク(代理)モデルが予測したイオン熱流束と元の輸送モデルが予測したイオン熱流束の関係を示した散布図。点が点線に近いほど予測精度が高い。



▲ 図2 最適化前(左)と最適化後(右)の代理モデルを用いて予測したイオン温度分布と元の輸送モデルで予測した分布の比較。



今号の表紙

ポスト「京」重点課題⑧ サブ課題C

準直接計算技術を活用したターボ機械設計・評価システムの研究開発

ターボ機械の設計高度化技術を構築するために、準第一原理的な手法である Large Eddy Simulation (LES) を用いた200角ファンの多目的最適化プロジェクトを実施している。(本文6ページ参照) 表紙は、重点課題⑧で開発中のFrontFlow/blue (FFB) によるファンの計算結果で、圧力のラプラシアン等の等値面を用いた翼面渦構造の可視化結果を中央に、翼負圧面における渦度の絶対値を左上に示す。最適設計の過程で得られる大量のデータをデータ科学的手法で分析することで、設計変数と性能の因果関係の理解が進むことが期待される。

東京大学生産技術研究所加藤千幸研究室 **三木 悠也**
(東京大学大学院工学系研究科機械工学専攻)

編集後記

近年注目される「AI・機械学習」は、HPCの利用法にも変革をもたらしつつあります。今回は、産業界及びアカデミアの最先端事例として、横浜ゴムのマテリアルズ・インフォマティクスによる革新的なゴム材料開発の取り組み、名古屋大学の機械学習による新素材合成の最適条件探索を高速化する取り組みを紹介するとともに、ポスト「京」重点課題における機械学習についても紹介します。



計算工学ナビ オフィシャルサイト

本誌のPDF版やソフトウェアライブラリ、ニュースなどのコンテンツを提供しているWebサイトは下記のURLからアクセスできます

<http://www.cenav.org/>



計算工学ナビ Vol.17

発行日：2019年9月19日

発行：東京大学生産技術研究所

革新的シミュレーション研究センター

〒153-8505 東京都目黒区駒場4-6-1

office@ciss.iis.u-tokyo.ac.jp