

機械学習のコモディティ化

田口 善弘 (中央大学理工学部 tag@granular.com)

1. はじめに

こんにちは、中央大学の田口です、と言ってももう、10年くらい物理学会で研究発表してないので多くの読者にとって「お前誰？」状態だと思うから簡単に自己紹介。今をときめく量子アニーリングの西森秀稔先生の(実質的な)最初の指導院生でしたが、量子系の研究はすぐにやめてしまい、そのあと、フラクタル、非線形物理学、粉粒体動力学(今で言うところのアクティブマター分野の源流みたいな分野です)をやったあと、今の大学に移動した1997年頃から徐々にバイオインフォマティクスに分野を移して、統数研の某Iさんには「田口さんもすっかり生物の人になったね」と言われて現在に至ります。バイオインフォというのは超ザックリいうと「機械学習で生物する」分野ですね。もっともアクティブなのは生物の中でも僕もやっているゲノム科学の大規模データの網羅的解析だと思います。

で、なんで生物学に移動したのってよく言われるんですが、僕自身は生物学に移動したつもりはなく、これは物理学だと思ってやっています。20年前に「これからはすぐに多変量解析の時代が来る、物理もそれを使ってやるようになる、そうなったら最大の課題は生物だ！」みたいな乗り²⁾だったんですが、まあ、「もうすぐ」が見事に20年かかりましたね(笑)。当時は機械学習って言葉は無かったので「多変量解析」と言っていました、やっとなんか手法的な部分は物理の一分野になりつつあります。で、生物学がやっぱり物理のターゲットと認識されるのはもうすぐかな。まあ、「もうすぐ」がまた20年かかったら自分が生きてるか怪しいわけですが。

2. 機械学習のこれから

シリーズ企画も始まったことだし、僕がここでいまさら「機械学習とは何か」みたいなことを語ってもしようがないので省略します。今、機械学習でなんかしようとするとなんかハードルは高いです。GAN(Generative Adversarial Network)とかすごく流行っていて、いろいろ画像の自動生成ができるようになり、人間が手でゼロから描いたのとはや区別不能なクオリティになりつつある。あ、面白そうやってみたい、と思ってもTensorFlowだのChainerなどの深層学習(Deep Learning, DL)のパッケージの使い方を勉強しないとイケないし、さらにどんなパラメータでどんなアーキテクチャでやったらいいのかわからないのでかなりの試行錯誤が必要です。なので「○○をDLでやりました」の○○が業界初だったらそれで論文になっちゃうレベ

ルです。

でもですね、これって結局、僕が駆け出しの院生のときに西森さん(あえてこう呼ばせていただきます)とやっていたことだってそうなんです。結局、やったことで新しいことは「スパコンを使って大規模疎行列を対角化する」。これだけです。大規模疎行列と言っても、当時使っていた東大のスパコンの最大メモリーが32「M」Bの時代ですからね。ギガバイトじゃなくて、メガバイト、ですよ。今時、ちょっと気の利いたスマホなら32GBのROMが乗ってるでしょ? いま、「大規模疎行列の対角化」ってスマホでできるんだと思います。30年強で技術はそこまでコモディティ化する。

いや、違うよ、行列の対角化は簡単だけど、DLはパラメータいっぱいだし、そんなメモリと計算速度が速くなくなっただけでできるようにならないでしょって言うなかれ。いま、ちょっと気の利いたパッケージを導入すれば、数値積分とか微分方程式の数値解とかほぼチューニング無しでできるよになっています。でもね、昔は違った。問題ごとに、収束しているかどうか、刻み幅はどうするか、とか考えないとできなかった。ガウス積分(誤差関数)を数値的に計算できなかったために卒研落とした学生だったわけですよ。でも、いま、誤差関数が入ってないパッケージなんてないですよ、メジャーな数値計算言語で、学生に「誤差関数計算できるプログラム書け」なんて誰も言わない。

そこまでじゃなくても、たとえば、エクセルでボタン一発で線形回帰とかできちゃう。多分、使っている人の大部分はどうやって線を引いているかさえ理解していないと思います。同じようにいまDLで最先端の「画像の中に壺があったら丸をつける」とか「アニメ風のかわいいキャラクターを描く」とか「顔認証でパスワードの代わりにする」みたいなことが、オフィスのプラグインでどんどん使えるようになっていく。実際、まだ性能は低いもののそういう自動的に機械学習を行うソフトは商用で作られて売られ始めています。

だから、今の機械学習がどんなにハードル高そうでも使いくさそうに見えても、プチッとボタン押したらできるようになります。要するにデータとラベルを用意して(まあ、ここが実はすごく大変なんです)アップロードして待っているとちゃんと学習済みのモジュールがパツと出る。2019年2月号に今田正俊先生たち³⁾が機械学習でスピン系の波動関数を求める話してましたけど、これもプチッと押したらできるようになる。すきなハミルトニアン組んで

投げると波動関数が帰ってくる。確実にそうなります。

3. 機械学習と物理学

そうすると、物理学はどうなるか。量子多体系とか統計力学とか、「難しい問題」が「機械学習で解ける難しい問題」の one of them に過ぎなくなる。物理学者がなんもしなくても、勝手に外側で機械学習が発展して、昨日まで解けなかったものがどんどん解けるようになって行く。これって物理にとって未体験ゾーンだと思うんですよね。物理は難しいから、他分野にわっと乗り込んで行って分野を席卷してその分野の昔の論文無視して引用しなかったりしてひんしゅくをかたりすることはよくあるわけですが、それを今度は逆にやられる。機械学習のすごい人が入ってきて、物理学者ポカーンのうちにどんどん問題を解いちゃう。そう、アルファ・スピンとかアルファ・ハバードとかで解いちゃうわけです。アルファなんとかってというのは話題になった囲碁などの完全ゲームを DL で解く一連のアプリケーションの接頭辞です。AlphaGO とかね。最近、AlphaFold とか銘打ってタンパクの立体構造予測にも殴り込んでかなりの成績をあげたので学術分野も実際に彼らのターゲットになっています。

問題が普遍化してしまったから、物理しかやってない物理学者は機械学習でなんでもアタックしている連中に勝てなくなる。まあ、国内市場しか相手してないと GAFA に歯が立たないのと同じようなことになる。「いや、機械学習じゃ答えは出せても理解はできない」って反論するかもしれない。でも、それも結構、危ういんですよね。僕はそこそこフォロワーが多い、非科学者も含んだツイッターアカウント持ってるんですけど、そこで「機械学習は物事を理解できない」みたいなツイートしたら 67 万件のインプレッション獲得して 2,000 回リツイートされて、4,000 回いいねされたんですが、そのレスに「結果を正しく予測できるのだからそれは新しい形の理解ではないか？」みたいなレスが、主に非科学者のフォロワーからたくさんついてびっくりしました。要するに、「何が理解か」っていう物理の牙城さえシフトしてしまう。下手すると物理学は哲学みたいに「なんの役にも立たない『理解』をこねくり回している変な学問」(哲学者の人ごめんさい！別に僕が本心からそう思っているわけじゃないです) な扱いになってしまいかもしれない、そういう未来さえ見えるわけです。

4. 物理学者が真に挑戦すべきこと

ここまですごく悲観的なこと書いたけど、実はこのピンチはチャンスだと思ってます。いままで物理学者は悪く言えばできることしかやってなかった。生物だって物質からできているから物質科学に過ぎないのに正面から向き合ってた。でも、今の機械学習はなんでも標的にでき

て、かつ、その基本はモンテカルロにせよ、ニューラルネットにせよ、かつて物理学者が生み出したり、発展に貢献したものがぐるっと回って戻ってきただけにすぎない。僕は仕事柄、バイオインフォや機械学習の会社の人に会うけど、僕が物理の教授だとわかるとみんな物理学者褒めちぎります。「ストリングやってた博士号持ち採ったけどすごく活躍してます！」みたいな。彼らいわく、統計や情報が出自の人は数学できるけど、現実の問題に興味が薄い。一方、実験を出た人は現実はわかるけど数学はだめ。現実の問題を数学で解こうという意欲は物理学者がピカイチですって。そうなんですよ。「青は藍より出でて藍より青し」がごとく、物理発の機械学習が立派になって戻ってきた。これを機会に物理学者はいままでやってこなかった全ての自然現象の解明に乗り出せばいいんだと思います。ランダウの理論物理学教程に入っていないものは物理じゃないなんてケチなことってはいけません。だって、もし、ランダウがまだ生きていたら、嬉々として「機械学習」をランダウの理論物理学教程の最後の一卷として付け加えるに違いないから。

5. おわりに

さて、一周回って我田引水して終わりにしたいと思います。いまだに生物の研究をしていると「なんで物理学科で生物の研究をしているんですか？」と学生さんにまで言われます。しかし、物理学はあくまで自然現象を数学的に解明するという科学である以上、生命がその対象外であっていいわけではないと思います。ラザフォードの逸話にみるまでもなく、生物学はいままで物理が挑戦するにはあまりにもデータが少なすぎました。しかし、今はゲノム科学の進歩のおかげで膨大なデータが解析できるようになっています。機械学習のような対象を限定しない手法が物理の標準的な方法の1つとして認識されるならば、生命科学のようないまだ物理の埒外に置かれがちだった研究対象も物理の一環として研究して行ってはいかがだろうか。若い人はなんてケチなことは言わない。老若男女全ての物理学会の会員がこれを機会にいままでのテーマにとらわれず、この世の全ては物理学、ぐらいの飛躍を期待しています。そして、生命科学についても、今とは逆に、物理学科に生命科学を研究している研究室が無かったら、学生が「なんでうちの物理学科には生命やっている先生いないんですか？」と文句を言うぐらいの時代が来てほしい、と願っています。以上！

参考文献

- 1) 坊農秀雅, 他, 『生命科学データ解析を支える情報技術』(技術評論社, 2019).
- 2) 田口善弘, 他, 『複雑性のキーワード』(共立出版, 2000) pp. 1-35.
- 3) 野村悠祐, 山地洋平, 今田正俊, 日本物理学会誌 74, 72 (2019).

(2019年2月28日原稿受付)