特集記事 シンポジウム報告

育種と AI—育種家は AI とどうつきあうか

二宮正士

東京大学大学院農学生命科学研究科, 東京都西東京市, 〒180-0002

Breeding and AI—How should breeders go further with AI?

Seishi Ninomiya

Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo 118-0002

キーワード

深層学習, 画像認識, 高速フェノタイピング, 育種戦略, 異分野融合

はじめに

近年, AI (Artificial Intelligence) という言葉を聞かない日は無い. 例えば, 画像内の対象物の判別問題で, 従来は対象物から判別に有効な特徴量を試行錯誤で抽出する必要があった. しかし深層学習はそのような特徴量抽出を不要にし判別モデル構築の姿を大きく変えた. また, 教師無し学習から類似データを新たに生成することや, 最適化問題の解決など, ここ 5 年ほど目を見張る成果を上げ, 人間の仕事を奪うと危惧する声も小さくない.

本シンポジウムは, 「育種での有効性はどうなのか?」, 「将来育種家にとって代わることができるのか?」といった, 学会員の素朴な疑問に関して議論を始めるために, AI の現状や課題について 4 名の方に話題提供をお願いした. 日本における AI 研究をリードしている情報科学系研究者 2 名に加え, 米国で実際に AI など先端情報科学を, 実践的にあるいは理論的に植物育種に活かそうとしている 2 名に話題提供をお願いした. ちなみにここでは, あえて AI を定義せず, 深層学習に代表されるような機械学習に関連する事項だけを対象にはしなかった.

演題 1 時空間データ解析のための最新 AI 技術

上田功功^{1,2} (1. NTT, 2. 理化学研究所革新知能統合研究センター)

日本における最新の AI 研究の例として, 時空間解析技術, 集団最適誘導技術, 地震被害推定技術について報告があった. 時空間解析技術では, カーネル回帰を用いて, 過去データが長期間用意できない場合でも時系列に沿っ

た空間パターンの将来変化を予測する手法を紹介した. これを用いて, イベント会場での人密度の空間変化がうまく予測できることが示され, 植生の空間的変化予測へ応用することの可能性を示唆した. また, 非格子状に設置された近傍の観測場所との距離・方向の情報から自己回帰するモデルを用いて, 空間上観測値が無い場所の観測値を予測する技術も紹介した. これでニューヨーク市において観測値が無い場所でのタクシー需要予測に成功した例を示した. この例も, 圃場での欠測株の観測値補完などに使えるかも知れない.

集団最適化技術では, 地震時の集団避難誘導のような場面で必要な人流ナビゲーション技術について紹介があった. 集団の誘導では, ①効率と安全のどちらを優先するのかといった評価指標, ②どれだけの人数をどの経路に配分するのかや, 輸送手段の増便方針などの誘導手段, ③グラフ理論を使うのか, セルオートマトンを使うのかといったシミュレーションモデル選択など, 問題設定のバリエーションは多様で, 誘導候補数が膨大なため網羅的シミュレーションが困難である.

このような技術が満たすべき要件として, 個人ではなく, 集団として最適な誘導を行う集団最適性, 現時点での状況 (観測) に応じた誘導ができるリアルタイム性, 渋滞が生じる前に予測に基づいた誘導ができる先行性がある. バイズ最適化と並列シミュレーションを用いることでそれらの課題を解決し, 例として, 大規模競技場への入退場人流最適化に成功した例を紹介した.

地震被害推定技術では, 南海トラフ地震を例に過去の地震を再現する摩擦パラメータを求め, シミュレーションで将来の地震を予測する問題を紹介した. 従来の人の仮説に基づく限定的な摩擦パラメータの探索からより網羅的な探索による学習と評価に切り替えることで, 過去地震の発生周期予測を大幅に改善できる例を示した.

2019 年 12 月 11 日受領

2020 年 6 月 3 日 J-STAGE 早期公開

Correspondence: snino@isas.a.u-tokyo.ac.jp

最後に、データ駆動型アプローチの限界について問題提起があり、今後はプロセスモデルとの融合が重要になると示唆した。また、第4の科学といわれるデータ集約科学 (Data Centric Science) の次として、シミュレーションデータから帰納的学習で真のモデルを推定し将来予測するような姿を一例としてあげた。

演題2 AIと画像解析

大倉史生^{1,2} (1. 阪大・産研, 2. JST さきがけ)

画像解析は特に深層学習と相性が良いことが知られ、畳み込みニューラルネットワーク (CNN) 等の応用が急速に進んでいる。その概要や植物への応用、技術的課題について報告があった。深層学習を用いることで従来困難であった物体認識 (対象が何であるかの判別) や画像分割 (必要な対象のみを画像から抽出) のための画像解析技術がこの10年で一気に発展している。たとえば、物体認識とその画像中の位置決めに関する大規模画像認識のコンテストである、ImageNet large scale visual recognition challenge (ILSVRC) では、2015年の提案手法が人間の認識精度を超えたと言われる。

最近では、画像から画像を再生成する深層学習研究も急速に発展し、雑音の多い画像をクリーンな画像にすることや、低解像度画像を高解像度に変換することなどが試みられている。また、同様の画像生成の一環として、画像中の同類の対象物 (例えば人) をまとめて抽出する Semantic Segmentation や同類の対象を個体別 (例えば個人別) に抽出する Instance Segmentation が盛んに研究されている。そのための基盤技術として教師なし生成モデル (generative model) の一種である GAN (Generative adversarial network, Goodfellow *et al.* 2014) への期待は高い。演者らは GAN の一種を応用して、葉のある植物の外観から枝構造を推定する研究を進めている (Isokane *et al.* 2018)。

深層学習は基本的にデータ駆動型アプローチであり、内挿は得意でも外挿能力にはまだ限界がある。これまで「ブラックボックス」であり、なぜ適切に機能するのか説明困難であったが、深層学習モデルの中身を理解しようとする試みも始まっている。判別のために画像のどこが着目されているのか、何が学習されているのか、といった点について深層学習の特定の層を解釈しようとするもので、演者らも葉の病徴判別で一例を示した (Toda *et al.* 2019)。今後、蓄積された生物学的知識と深層学習のようなブラックボックスの融合が重要であるとした。

演題3 Case example of the use of high-throughput phenotyping and AI for plant breeding

Asheesh Singh (Iowa State Univ.)

実際にダイズ育種家として品種育成を行う中でどのように高速育種法やAIを活用しようとしているのか自らの実践例も含めて報告した。現在、①交配時期の最適化、

②交配親の選択と最適化、③観測空間や観測スペクトルの最適化、④フェノタイピング手法の開発、⑤複数形質の同時最適化、⑥深層学習による収量予測、⑦最適なゲノミックセレクション (GS) 手法などについて研究開発を進め、育種パイプラインの中に取り組むことを進めつつある。その中で開発された畦間を自動走行し成熟期のダイズを横から撮影して鞘数を自動計数するためのロボット (Gao *et al.* 2018)、スマートフォンでダイズの生理障害程度などを計測するツール (Naik *et al.* 2017)、深層学習におけるネットワーク層の反応を理解することで多数のダイズ病害種の判別ばかりでなく病害程度も診断可能としたモデル (Ghosal *et al.* 2018)、ハイパースペクトル画像の波長選択を遺伝アルゴリズムと機械学習の組合せで最適化し、ダイズ炭腐病の早期発見を可能にしたモデル (Nagasubramanian *et al.* 2018)、深層学習によるセンチュウ卵の認識と計数モデル (Akintayo *et al.* 2018) など、一部は実際のダイズ育種現場にも導入されている高速フェノタイピング技術を紹介した。

本講演では、高速フェノタイピング技術開発やAIを多用する新しい育種技術開発において、異分野連携と異分野連携でリーダーシップをとれる人材育成の重要性が強調された。アイオワ州立大学では、学内の植物科学研究所 (Plant Science Institute) に、育種学、植物病理学、ロボット工学、情報科学などの専門家が結集する枠組みがあり、そこでのつながりが次世代育種技術開発や人材育成に大きく貢献していることが紹介された。

演題4 Improving genomic selection with optimization and machine learning

Lizhi Wang (Iowa State Univ.)

ゲノミックセレクション (GS) による育種親選択の改善に関する提案手法 LAS (Look-Ahead Selection, Moeiniazade *et al.* 2019) を中心に報告した。一般に GS による交配親の選択には短期的な遺伝的獲得量と長期的な遺伝的改良能力とのバランスをどのように取るかという課題がある。これまでの GS には CGS (Conventional Genomics Selection), OHV (Optimal Haploid Value, Daetwyler *et al.* 2015), OPV (Optimal Population Value, Goiffon *et al.* 2017) などが提案されている。CGS は、ゲノミック推定育種価 (GEBV) を元に最良の個体を選抜すること、即ち短期的な獲得に主眼を置いている。一方、OHV は優れた倍化半数体を後代に作出できる能力に着目して個体選抜するが、これはより後代の遺伝的改良能力に着目して選抜していることになる。OHV を拡張した OPV は、個体の選抜ではなく、世代を重ねたときに最良の後代を作出する個体群を選択することを目指している。

以上の三つの GS 手法では、①いつまでに育種目標を達成するかという時間、②どのような交配戦略をとるか、③育種リソース (交配数や維持する後代の数など) が考慮されていない。LAS は、実際の GS で制約となる以上

の3点を考慮した上で、最適の交配親を決定するアルゴリズムとして提案された。LASでは、「先読み」情報を導入し、限られたリソースで特定の時間までに高い遺伝的獲得量を達成する確率を評価する。LASの目的関数は最終世代における最良後代のGEBV期待値として定義し、選択された育種親、交配方法、組換え頻度、および残りの世代数を入力に含む。前の2つのインプット項はモデルによって最適化される必要がある変数で、後者の2つは、最適なソリューションを探索する際にモデルが考慮する必要があるパラメータである。LASをトウモロコシ近交系データを用いたシミュレーションでCGSやOHV、OPVと比較したところ、LASは、①育種年限において最も高い遺伝的獲得量を得ること、②育種年限の最後まで最高レベルの遺伝的多様性を保持していること、③1000回のシミュレーションの中のほとんどで、他の手法に比べ最も高い遺伝的獲得量を得ることが示された。

総合討論

講演後の総合討論では、各演者への質問に加え、以下の点について議論が交わされた。

・外挿の問題

2名の講演者が、機械学習のアプローチ、即ち単純なデータ駆動型アプローチの問題として、その「外挿性の限界」を指摘した。これは実際に観測された学習データの範囲内ではとても有益であるが、未知の空間の予測や判別は原理的にも苦手であるということである。例えば、より高い収量を予測するといった育種の場面でありがちな状況にうまく使えないかもしれないことを示唆している。データに依存する既存のAIの弱点を補完して問題を解決するために、既存の知識に基づくプロセスモデルとの融合など新たな研究展開の必要性が強調された。

・学習コストの軽減

深層学習など機械学習モデルの精度は、学習データの品質に極めて高く依存している。その品質は、十分なデータ数に加え、アノテーション（例えば物体認識であればそれが何かというラベル付け）の正しさに決まるが、通常アノテーションは手作業で膨大な時間コストを要し、しばしばボトルネックになっている。

データオーギュメンテーション、生成モデルなどによる学習データの水増し、遷移学習等に加え、少ない学習データで構築したモデルの誤判別データの再学習により学習時間コスト低減が実現することが報告された。

また、十分な学習データとなるビッグデータがそもそも育種研究用として存在するののかという点も議論になったが、大量に保存されている歴史的データを活用しきることが重要で、そのためには、時に手書きで残された紙媒体のデータも含むアナログデータの効率的なデジタル化や、不斉一で欠損が多いデータの補完手法、メタアナリシスなどに関する研究開発の重要性が確認された。

・AIから新知識の獲得

これまで深層学習に代表されるAIは、通常その中身を理解することは難しいとされたが、最近はCNNの特定層ヒートマップの解析などを通して、新たな応用が始まっている。今後、うまく機能するAIから新知識を導出することにも大きな可能性があることも指摘された。

・分野間連携の課題

AI分野は、急速に発展する分野であるが、極めて強力なツールとして育種分野でも高く期待されている。一方、それを十分に使いこなすことは通常の育種研究者にとって容易ではない。アイオワ州立大学は、農学、工学、情報科学など異分野の研究者が所属する組織を立ち上げ、毎週定期的に議論を行う場を設けることで、課題解決に向けた学際連携体制を作り上げている。それを単に共同研究ばかりでなく、学生の教育にも最大限活用している取り組みは、とても示唆に富んでいた。

おわりに

「AIが育種家」に取って代わるかという問いかけに対して、全ての演者が否定的であった。それは現在のAIが外挿に弱いことに示されるように、100%データ駆動だけでは必ずしも「クリエイティブ」になれない点にある。一方、組合せ数が多い非常に複雑な最適化問題では、人間の能力を遙かに凌ぐことや、AIを駆使した高速フェノタイピング技術が、明らかに育種の効率化を促進する例も示され、うまくAIを取り込んで既存技術と相互に補完し合うのが当面の最適解ということになった。

本シンポジウムは、JST CREST「フィールドセンシング時系列データを主体とした農業ビッグデータの構築と新知見の発見」(JPMJCR1512)および、JST さきがけ「情報協働栽培」(https://www.jst.go.jp/kisoken/presto/research_area/ongoing/bunyah27-6.html)との共催で開催された。

引用文献

- Akintayo, A., G. Tylka, A. Singh, B. Ganapathysubramanian, A. Singh and S. Sarkar (2018) A deep learning framework to discern and count microscopic nematode eggs. *Sci. Rep.* 8: 9145.
- Daetwyler, H.D., M.J. Hayden, G.C. Spangenberg and B.J. Hayes (2015) Selection on optimal haploid value increases genetic gain and preserves more genetic diversity relative to genomic selection. *Genetics* 200: 1341–1348.
- Gao, T., H. Emadi, H. Saha, J. Zhang, A. Lofquist, A.K. Singh, B. Ganapathysubramanian, S. Sarker, A.K. Singh and S. Bhattacharya (2018) A novel multirobot system for plant phenotyping. *Robotics* 7: 61.
- Ghosal, S., D. Blystone, A.K. Singh, B. Ganapathysubramanian, A. Singh and S. Sarker (2018) An explainable deep machine vision framework for plant stress phenotyping. *Proc. Natl. Acad. Sci. USA* 115: 4613–4618.
- Goiffon, M., A. Kusmec, L. Wang, G. Hu and P. Schnable (2017) Improving response in genomic selection with a population-

- based selection strategy: optimal population value selection. *Genetics* 206: 1675–1682.
- Goodfellow, I., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio (2014) Generative adversarial networks. arXiv:1406.2661.
- Isokane, T., F. Okura, A. Ide, Y. Matsushita and Y. Yagi (2018) Probabilistic plant modeling via multi-view image-to-image translation. Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR'18) arXiv:1804.09404.
- Moeiniazade, M., G. Hu, L. Wang and P.S. Schnable (2019) Optimizing selection and mating in genomic selection with a look-ahead approach: An operations research framework. *G3* 9: 2123–2133.
- Nagasubramanian, K., S. Jones, S. Sarkar, A. Singh, A. Singh and B. Ganapathysubramanian (2018) Hyperspectral band selection using genetic algorithm and support vector machines for early identification of charcoal rot disease in soybean stems. *Plant Methods* 14: 86.
- Naik, H.S., J. Zhang, A. Lofquist, T. Assefa, S. Sarker, D. Ackerman, A. Shingh, A. Singh and B. Ganapathysubramanian (2017) A real-time phenotyping framework using machine learning for plant stress severity rating in soybean. *Plant Methods* 13: 23.
- Toda, Y. and F. Okura (2019) How convolutional neural networks diagnose plant disease. *Plant Phenomics* 2019: ID 9237136.

当日 演題

はじめに 二宮正士（東京大学大学院農学生命科学研究科）

講演

1. 時空間データ解析のための最新 AI 技術
上田修功^{1,2}（1. NTT, 2. 理化学研究所革新知能統合研究センター）
2. AI と画像解析
大倉史生^{1,2}（1. 阪大・産研, 2. JST さきがけ）
3. Case example of the use of high-throughput phenotyping and AI for plant breeding
Asheesh Singh（Iowa State Univ.）
4. Improving genomic selection with optimization and machine learning
Lizhi Wang（Iowa State Univ.）

総合討論