

令和 4 年 5 月 6 日現在

機関番号：11201

研究種目：基盤研究(A) (一般)

研究期間：2019～2021

課題番号：19H00938

研究課題名(和文) イネの野外ビッグデータを用いた「人工知能」による育種プラットフォームの開発

研究課題名(英文) Artificial Intelligence Assistance for Plant Breeding

研究代表者

下野 裕之 (Shimono, Hiroyuki)

岩手大学・農学部・教授

研究者番号：70451490

交付決定額(研究期間全体)：(直接経費) 35,200,000円

研究成果の概要(和文)：「人工知能」の育種への利用により品種育成を補助するシステムを開発を目指した。まず、(1)「表現型」の標準化として、イネ収量の野外ビッグデータから、表現型である品種特性を作物モデルを用いて環境に対する回帰係数として抽出するプラットフォームの開発した。続いて、(2)「遺伝子型」の標準化として、163万箇所のSNPデータに基づく365品種/系統の遺伝子型プラットフォームを構築した。また、(3)「人工知能」での学習として、「シングル・エポック・ラーニング法」を提案した。最後に(4)既存のゲノミック予測法およびその作物モデルとの融合による基準精度の評価を行った。

研究成果の学術的意義や社会的意義

地球温暖化により気候変動が拡大する中、その変化に適応できる新品種の育成が喫緊の課題である。本研究では、イネをモデル植物として、過去に蓄積された野外ビッグデータを再利用する新たな整理方法を提案するとともに、その表現型情報と遺伝子型情報について人工知能を用いてはじめて解析した試みに位置付けられる。今後は、イネ以外のダイズやコムギなど他の作物への応用とともに、さらに発展的に蓄積されたデータを利用し、情報を再解析することで農業生産上に有用な遺伝子領域の特定や、将来の気象条件に合わせた新品種の開発に寄与することが期待される。

研究成果の概要(英文)：We aimed to develop a system that assists breeding by using "artificial intelligence" for breeding. First, (1) as a standardization of "phenotype", we developed a platform to extract phenotypic cultivar characteristics as a regression coefficient for the environment from field big data of rice yield using a crop growth model. Subsequently, (2) as a standardization of "genotype", we constructed a genotype platform of 365 varieties / strains based on SNP data of 1.63 million locations. We also proposed (3) "Single epoch learning method" as learning with "artificial intelligence". Finally, (4) the standard accuracy was evaluated by fusing with the existing genomic prediction method and its crop model.

研究分野：作物学

キーワード：人工知能 育種 多収 ゲノム 多型 成長モデル 水稻

1. 研究開始当初の背景

植物育種は、作物の収量性と品質の向上を通じて食料生産に大きく貢献してきた。地球規模で進行している気候変動に適応するには革新的な育種法の確立が喫緊の課題である。本研究は、「人工知能」の育種への利用により品種育成を補助するシステムの構築を目的とした。必要となるイネの収量など重要な農業形質の「表現型」は野外ビッグデータとして全国の農業試験場で長期にわたり蓄積されている。しかし、野外かつ群落条件で取得される農業形質データは、栽培時の気象の影響を強く受け一般化が難しい。また「育成情報」や「遺伝子型」を「人工知能」で統一的に評価する手法、さらにニューラルネットワークのノード間を結ぶ最適な関数型などの知見がない。

2. 研究の目的

本研究では、「人工知能」による汎用性の高い育種プラットフォームを、(1)「表現型」の標準化、(2)「遺伝子型」の標準化、(3)「人工知能」での学習、(4)既存のゲノミック予測法およびその作物モデルとの融合による基準精度の評価を通じ、イネの野外ビッグデータを用いた「人工知能」による育種プラットフォームを開発した。

3. 研究の方法

(1)「表現型」の標準化：水稻奨励品種決定基本調査成績データベース(2018/12版)(奨励データ)に掲載されている収量など26項目を1セットする207,331データセット(8025品種・系統×38年×110地点)について異常値を精査し用いるとともに、年次・地点の日々の気象データ(農研機構・モデル結合型作物気象データベース)をあわせてAccess(マイクロソフト社)に収納した。各データセットに該当する生育期間別の気象の平均値など25項目ならびに潜在収量等の作物モデル(Masuya and Shimono, 2017)からの出力値12項目を算出した。

(2)「遺伝子型」の標準化：イネ品種・特性データベース(農研機構)に記載されている品種/系統の系譜情報を整理したネットワークおよび*k*-medoids法による解析から、代表的な200品種/系統を選定した。また、ネットワークのエッジ、すなわち祖先品種も選定した。全国の水稲品種開発機関および農研機構ジーンバンクから種子の分譲を受けた294品種/系統のDNAを抽出し、次世代シーケンサーでシーケンスした。加えて、公共アーカイブに保存されている71品種/系統のシーケンスリードも利用した。「日本晴」リファレンス(IRGSP1.0)にリードをアラインメントし、SNP情報を得て、imputation(BEAGLE5.1)を実行し、計365品種/系統の遺伝子型を決定した。表現型プラットフォームから得られた品種/系統の特性値を用いて、線形混合モデルによるGWAS(R package 'gaston')を行った。

(3)「人工知能」での学習：シングル・エポック・ラーニング(SEL)を用いた作物の遺伝形質予測の自動化と高速化のため、新型のニューラルネットワーク(NN)訓練アルゴリズムを開発した。ディープラーニングに代表されるNNを用いたデータ解析は、ネットワーク構造の調整、活性化関数の調整、学習率の調整など専門知識が必要な手作業による調整を自動化・省略化することのできるニューラルネットワークの訓練方法を提案した。提案法を作物遺伝形質予測に応用し、有効性を確認した。

して選定された品種の内訳をみると、わが国の育種において多く交配等に利用されてきた品種/系統と考えられた。最も後代の数が多い品種は「コシヒカリ」であり、2番目は「ひとめぼれ」、3番目は「日本晴」であった。得られたシーケンスリードから、およそ163万箇所のSNPデータに基づく365品種/系統の遺伝子型プラットフォームを構築した。この遺伝子型情報を用いたSTRUCTURE解析(sNMFプログラム)では、365品種/系統はindica等品種、陸稲品種、北海道品種および本州等品種の4つの遺伝的クラスターに分けられた。表現型プラットフォームから出力した「穂数400本のときの全重」を形質値としたGWASでは、形質に関与すると考えられる複数のゲノム領域を検出した。

(3)「人工知能」での学習：UC Irvine Machine Learning Repositoryより取得したデータセットにSELを適用したところ、BPと比較して学習誤差の縮小と学習時間の短縮が確認できた。また、イネの奨決データに応用したところ、主要10品種の収量の形質に対し予測誤差2.9%の結果が得られた。一般的な解析手法であるBLUPと比較すると、開発したSELは誤差を47%低減した。線形モデルであるBLUPと比較して非線形モデルであるNNは入力データのより複雑な特徴を捉えて予測誤差を低減していると考えられた。本方式は、ネットワーク構造の調整、活性化関数の調整、学習率の調整などを行うことが難しい専門知識を持たないユーザでも、学習に用いるデータを準備するだけで利用できるため、イネ遺伝形質予測プラットフォームとして幅広いユーザによる活用が期待できる。

(4)既存のゲノミック予測法およびその作物モデルとの融合による基準精度の評価：奨決データの表現型情報と、品種/系統の系譜情報、気象情報、ゲノム情報を組み合わせて、ゲノムと表現型の関係をモデル化するため、遺伝子型値のBLUPより、品種の配布年にとともなう変化、育成地間の違いを確認することができた。ゲノム情報(実現血縁行列)のみを用いた予測では、系譜情報(分子血縁行列)を用いた予測に比べ、精度が高かった。しかし、ゲノム情報の無い品種/系統の遺伝子型値も用いて分子血縁行列をもとに予測を行った場合、大幅に予測精度が向上した。また、いずれの形質も、実現血縁行列と分子血縁行列を混合した血縁行列を用いて予測を行なった場合に最も精度が高かった。以上のように、これまでに蓄積された育種データを系譜、ゲノム、気象データと併せて解析することにより、有用な情報を引き出すことができることが示唆された。

引用文献

- Christensen OF, Madsen P, Nielsen B, Ostersen T, Su G. 2012.** Single-step methods for genomic evaluation in pigs. *Animal* **6**(10): 1565-1571.
- Legarra A, Aguilar I, Misztal I. 2009.** A relationship matrix including full pedigree and genomic information. *Journal of Dairy Science* **92**(9): 4656-4663.
- Masuya Y, Shimono H. 2017.** Mining a yield-trial database to identify high-yielding cultivars by simulation modeling: A case study for rice. *Journal of Agricultural Meteorology* **73**(2): 51-58.

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計10件（うち招待講演 6件 / うち国際学会 0件）

1. 発表者名 小山田絢子・佐藤直之・佐藤睦志・下野裕之
2. 発表標題 イネ表現型ビッグデータを用いた品種特性の抽出法. I. アプリの開発
3. 学会等名 第249回日本作物学会
4. 発表年 2020年

1. 発表者名 佐藤直之・小山田絢子・佐藤睦志・下野裕之
2. 発表標題 イネ表現型ビッグデータを用いた品種特性の抽出法. . アプリによる解析
3. 学会等名 第249回日本作物学会
4. 発表年 2020年

1. 発表者名 陳泰伸・佐藤睦志・阿部陽・山崎将紀・下野裕之・岩田洋佳
2. 発表標題 作物成長モデルに基づいて構築した環境カーネルを用いたイネ表現型の予測精度向上に関する研究
3. 学会等名 第139回講演会日本育種学会
4. 発表年 2021年

1. 発表者名 岩田洋佳・佐藤睦志・阿部陽・金天海・師田郷太・下野裕之
2. 発表標題 血縁情報とゲノム情報を用いたイネ奨励品種決定試験データの遺伝解析
3. 学会等名 第141回講演会日本育種学会
4. 発表年 2022年

1. 発表者名 下野裕之・佐藤睦志
2. 発表標題 表現型プラットフォームの開発：成長モデルを用いた作物収量を定量化する新たな評価法
3. 学会等名 第253回日本作物学会（招待講演）
4. 発表年 2022年

1. 発表者名 阿部陽・佐藤睦志・岩田洋佳・金天海・下野裕之
2. 発表標題 遺伝子型プラットフォームの開発：GWASによる新たな収量性遺伝子の探索
3. 学会等名 第253回日本作物学会（招待講演）
4. 発表年 2022年

1. 発表者名 金天海・岩田洋佳・阿部陽・佐藤睦志・下野裕之
2. 発表標題 シングル・エポック・ラーニング ディープラーニングを用いた作物の遺伝形質予測の自動化と高速化
3. 学会等名 第253回日本作物学会（招待講演）
4. 発表年 2022年

1. 発表者名 岩田洋佳・佐藤睦志・阿部陽・金天海・師田郷太・下野裕之
2. 発表標題 これまでに蓄積されてきたイネ育種データをゲノム-表現型関係モデリングに活用する
3. 学会等名 第253回日本作物学会（招待講演）
4. 発表年 2022年

1. 発表者名 Iwata, H.
2. 発表標題 How to model GxE and use it for plant breeding: examples from our researches.
3. 学会等名 UNL Agronomy Seminar Series, Nebraska Lincoln University, NE, USA (招待講演)
4. 発表年 2021年

1. 発表者名 Iwate, H.
2. 発表標題 Modeling Environmental Responses of Crops: Using Historical Data of Rice and Field Phenomics Data of Soybean.
3. 学会等名 ASA, CSSA, SSSA International Annual Meeting, Salt Lake City, UT, USA (招待講演)
4. 発表年 2021年

〔図書〕 計0件

〔出願〕 計1件

産業財産権の名称 アクセスコード	発明者 金天海	権利者 同左
産業財産権の種類、番号 特許、PCT/JP2021/46271	出願年 2021年	国内・外国の別 国内

〔取得〕 計0件

〔その他〕

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	岩田 洋佳 (Iwata Hiroyoshi) (00355489)	東京大学・大学院農学生命科学研究科(農学部)・准教授 (12601)	
研究分担者	阿部 陽 (Abe Akira) (80503606)	公益財団法人岩手生物工学研究センター・ゲノム育種研究部・主席研究員 (81202)	

6. 研究組織（つづき）

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分 担 者	金 天海 (Kim Chyon) (30424815)	岩手大学・理工学部・准教授 (11201)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関