

令和 6 年 5 月 21 日現在

機関番号：12601

研究種目：基盤研究(C)（一般）

研究期間：2021～2023

課題番号：21K06354

研究課題名（和文）大規模な植物形質データベースと機械学習の活用による形質間関係の網羅的・定量的評価

研究課題名（英文）Exhaustive analyses on interrelationships among plant functional traits based on application of machine learning to an open access database

研究代表者

饗庭 正寛 (Aiba, Masahiro)

東京大学・大学院農学生命科学研究科（農学部）・特任助教

研究者番号：80751990

交付決定額（研究期間全体）：（直接経費） 3,100,000円

研究成果の概要（和文）：植物の機能形質データベースTRYに登録された1875種の10形質を対象に機械学習を活用した形質間関係の解析を行った。全ての形質について機械学習モデルの予測能力が単回帰モデルの予測能力を上回っており、形質間関係における交互作用や非線形性の重要性が示唆された。またFriedman's Hの算出により、形質間関係で説明されるばらつきが最大30%程度が形質間の交互作用によるものであることが実証された。さらに国産樹木157種の18形質を対象に形質の多次元空間内における配置を再現するために必要な形質数を距離相関により検証した。その結果、距離相関0.9を達成するためには最低6形質が必要なが示された。

研究成果の学術的意義や社会的意義

生物の形質間関係は、基礎科学的に重要なだけでなく、気候変動や土地利用変化といった人為攪乱が生態系の機能やサービスに与える影響を予測・評価するうえでも不可欠な情報である。本研究で発見された植物の形質間関係の様々な特性はそうした予測・評価に係る応用研究の精度向上に寄与するものである。また、形質多次元空間内の配置に関する知見は応用研究においても重要性の高い機能的多様性の効率的な活用に繋がる可能性が高い。

研究成果の概要（英文）：Interrelationships among functional traits were analyzed by using a machine-learning method for 10 traits of 1875 species available in the TRY database. For all traits, the predictive ability of the machine learning model exceeded that of the single regression model, suggesting the importance of relationships among multiple traits, interactive effects, and nonlinearities in the relationships among traits. The calculation of the Friedman's H index demonstrated that up to 30% of the explained variation was attributable to interactions among traits. In addition, the number of traits required to reproduce the arrangement of traits in the multidimensional trait space was verified by using distance correlation for 18 traits of 157 Japanese tree species. The results showed that at least six traits are required to achieve a distance correlation of 0.9.

研究分野：生態学

キーワード：機能形質 機能的多様性 植物生態学 森林生態学

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

1. 研究開始当初の背景

植物の機能形質間の関係は、植物の機能・生活史の多様性の基盤であり、その全容の解明は様々な環境問題とも深く関わる重要な課題である。近年の機能形質データの飛躍的な充実の結果、葉経済スペクトル(多数の葉形質間の強い相関関係)の発見に代表されるように、形質間関係に関する研究は大きな進展を遂げてきた。しかし、形質間の重要な関係が線形な相関関係に限定されると考える合理的な理由はないにもかかわらず、形質間の交互作用や非線形な関係を検証した例はほとんどなかった。形質間関係における交互作用や非線形性の重要性を検証することは気候変動や土地利用変化といった人為攪乱が生態系の機能やサービスに与える影響の予測精度向上のためにも重要であった。

2. 研究の目的

本研究では近年公開された世界最大の植物形質データベース TRY のデータを、機械学習法を用いて解析することで、形質間関係において、交互作用・非線形性は定量的にどの程度重要か? 形質間関係は、気候、地理区分、系統などのコンテキストに依存してどう変化するか? 植物の形質多様性は何次元か? また、その次元数はコンテキストに依存して変化するか? の 3 つの問いに答えることを目的とした。

3. 研究の方法

欠測値の数に基づく絞り込みの結果、機能形質データベース TRY に登録されたデータのうち、1875 種を対象とし、説明変数としてはカテゴリカルな形質を含む 12 形質、応答変数としてはカテゴリカルな形質を除く 10 形質を用いた。単相関が最大の形質を用いた単回帰モデルと全形質を用いた勾配ブースティング回帰木による機械学習モデルの予測性能を交差検証により比較することで形質間関係における交互作用や非線形性の重要性を定量評価した。また、無作為化後の予測性能低下度に基づいて各説明変数の重要度を評価した。その結果、特定された特に重要な形質間関係について Partial dependence plot を用いて可視化することによる非線形性の評価も行った。さらに Friedman's H の算出により形質間関係における交互作用の重要性を定量評価した。また説明変数に系統情報を追加してこれら一連の解析を再度行うことで形質間関係の系統依存性を検証した。最後に、欠測値の少ない国産樹木 157 種の 18 形質を対象に形質の多次元空間内における配置を再現するために必要な形質数を距離相関により検証した。

4. 研究成果

すべての形質について、機械学習モデルの予測能力が単回帰モデルの予測能力を上回っており、形質間関係における複数形質間の関係、交互作用、非線形性の重要性が示唆された(Fig. 1)。草丈・樹高(単回帰の決定係数 0.41、機械学習の決定係数 0.82)、茎・材密度(単回帰 0.38、機械学習 0.67)、重量あたり葉面積(単回帰 0.21、機械学習 0.61)、葉乾重密度(単回帰 0.09、機械学習 0.56)といった形質で特に機械学習モデルの効果が大きかった。一方、種子重量(単回帰 0.47、機械学習 0.58)や葉重量あたり炭素含量(単回帰 0.09、機械学習 0.15)などの形質では、機械学習モデルの優越性は限定的であった。機械学習モデルにおいて、説明変数として最も重要度の高い形質は最も単相関の強い形質と必ずしも一致しなかった。たとえば草丈・樹高では、最も単相関の強い形質は種子サイズだったが、機械学習モデルにおける重要度は茎・材密度と葉サイズの重要度が高く、種子サイズの重要度は 3 番目であった。葉乾重密度では、最も単相関の強い形質は茎・材密度だったが、機械学習モデルにおいては重量あたり葉面積、草丈・樹高、葉の厚さ等の重要度が高く、茎・材密度の重要度は 6 番目であった。重要な形質間関係を Partial dependence

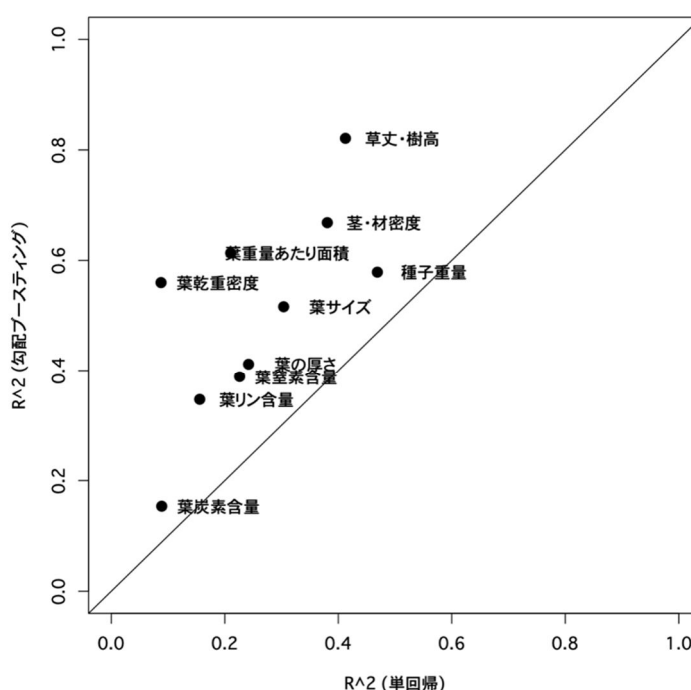


Fig.1 単回帰モデルと機械学習モデルの予測性能の比較

plot により可視化することで、形質間関係の非線形性を視覚的にも示した。たとえば草丈・樹高が 1m 以上の種においては草丈・樹高と種子重量の間に正の相関が見られたが草丈・樹高 1m 未満の種においてはそのような傾向は見られなかった。また、重量あたり葉面積が小さい種においては葉乾重密度との間に負の相関が見られたが重量あたり葉面積が小さい種ではそのような傾向は見られなかった。Friedman's H を用いた検証の結果、今回対象とした形質間関係においては最大で形質間関係で説明されるばらつきの 30%程度が他の形質の値に依存したものであることが示された。たとえば、葉の重量あたりリン含量と葉の重量あたり炭素含量の間の関係は葉の重量あたり窒素含量に依存して異なっていた。

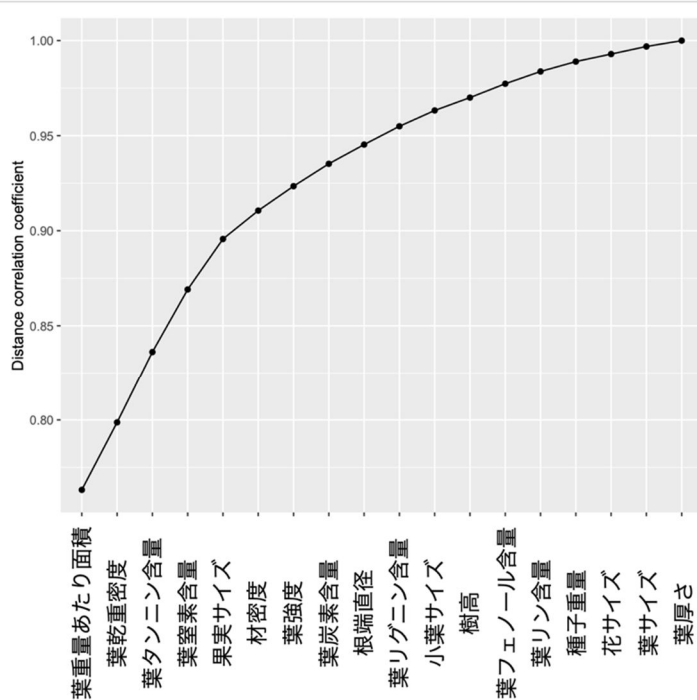


Fig.2 増分の大きい順に形質を追加した場合の全形質との距離相関の推移

次に形質予測モデルに系統情報を追加し、系統情報のみを説明変数とするモデルと系統情報と形質の両方を用いたモデルを比較すると両方を用いたモデルの予測性能が大幅に高く、近縁種間においても形質間関係が重要であることが示された。また形質のみを説明変数とするモデルと系統情報と形質の両方を用いたモデルを比較すると両方を用いたモデルの予測性能がやや高くなっており今回のモデルに含まれずかつ系統的に保存された形質の重要性もしくは系統と形質の交互作用の存在が示唆された。Friedman's H の算出により 2 つの形質間関係に対する系統の影響を検証すると一部の形質間関係ではばらつきの 20%程度が系統に依存していた。例えば、草丈・樹高と葉のサイズの間の正の相関は裸子植物では見られなかった。

さらに植物の機能形質多次元空間の有効な次元数についての解析に取り組んだ。これまでの解析に引き続き、植物の機能形質データベース TRY に登録されたデータのうち、1875 種の 10 形質を対象とした解析を試みたが欠損値の影響が大きかった。そのため本解析については代表者が以前に収集した国産樹木 157 種の 18 形質を対象に行うこととした。いくつかの手法を検討した結果、多変数同士の非線形な相関を検証する手法である距離相関に基づく評価を行った。まず全 18 形質を用いた時の種間距離との相関が最も強い形質を総当りで検出し、以降も距離相関の増分が大きい順に形質を追加した。全 157 種を対象とした場合、距離相関 0.9 を達成するためには最低 6 形質が必要であり 0.95 を達成するためには、最低 10 形質が必要であった (Fig. 2)。重要な形質は追加順に重量あたり葉面積、葉乾重密度、葉のタンニン含量、葉の窒素含量、果実サイズ、材密度であった。被子植物のみ、常緑樹のみ、落葉樹のみといったカテゴリー別に解析すると重要な形質はある程度変動した。樹木の研究で使用頻度の高い樹高、重量あたり葉面積、種子重量の 3 形質と全 18 形質の距離相関は 0.80、材密度と葉サイズを加えた 5 形質との距離相関は 0.84 であり、これらの形質を選択することが最善ではない可能性が示唆された。

これまで線形な相関関係の解析が大半を占めていた植物の機能形質間関係における、形質間関係における複数形質間の関係、交互作用、非線形性の重要性やそうした関係の系統依存性、形質多次元空間内の配置を決定する形質の数や種類に関する本研究の発見は、気候変動・人為攪乱が形質間関係を介して生態系機能に与える影響の予測性向上や機能的多様性評価における形質選択のガイドラインの提示など、応用研究においても幅広い活用が期待される。

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件/うち国際共著 0件/うちオープンアクセス 1件）

1. 著者名 Masahiro Aiba, Rei Shibata, Michio Oguro, Tohru Nakashizuka	4. 巻 13
2. 論文標題 Variable effects of vegetation characteristics on a recreation service depending on natural and social environment	5. 発行年 2023年
3. 雑誌名 Scientific Reports	6. 最初と最後の頁 684
掲載論文のDOI（デジタルオブジェクト識別子） 10.1038/s41598-023-27799-7	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計3件（うち招待講演 0件/うち国際学会 0件）

1. 発表者名 梶野 浩史, 小黒 芳生, 黒川 紘子, 小野田 雄介, 饗庭 正寛, 彦坂 幸毅
2. 発表標題 日本国内の森林の地上部バイオマスと気候条件、樹木機能形質の関係
3. 学会等名 第134回日本森林学会大会
4. 発表年 2023年

1. 発表者名 饗庭正寛, 小黒芳生, 柴田嶺
2. 発表標題 陸域生態系サービスの全国評価とシナリオ分析
3. 学会等名 日本生態学会第69回全国大会
4. 発表年 2022年

1. 発表者名 Hirofumi KAJINO, Michio OGURO, Hiroko KUROKAWA, Yusuke ONODA, Masahiro AIBA, Kouki HIKOSAKA
2. 発表標題 Relationships of above-ground biomass with climate and species composition across Japanese natural forests: which types of forests stock more carbon?
3. 学会等名 日本生態学会第71回全国大会
4. 発表年 2024年

〔図書〕 計0件

〔産業財産権〕

〔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
--	---------------------------	-----------------------	----

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
---------	---------