

データ駆動モデルによる生物の空間分布解析 Data-driven model for species distribution analysis

○福田信二¹
○FUKUDA Shinji

1. はじめに

生物の空間分布や移動分散とその支配要因については古くから盛んに研究がなされている。それらの多くは生態学的な基礎研究であったが、人間活動と環境との調和への認識の広まりに伴い、工学分野との学際研究としての取り組みが増加している。さらに、近年では、生物多様性情報の大規模データベース化とオープンソフトによる解析技術の汎用化の世界的な進展により、機械学習を用いたデータ駆動モデルの生態学分野での応用研究は、エコインフォマティクス (Ecoinformatics, Ecological informatics) の一分野として、急激な成長を遂げている。本報では、福田 (2015) に基づいて、空間分布モデルを事例に、データ駆動モデルの応用研究について紹介する。

2. エコインフォマティクスとデータ駆動モデル

エコインフォマティクスは、生態学分野における情報科学の応用研究であり、生態データの収集から解析や可視化までの大部分が対象となる。解析面では、人工知能技術の適用研究から大きな展開が始まり、近年の機械学習の進歩により研究展開が加速し続けている。とくに、生物・生態データの電子化や各種センシング技術の発達などにより、生態学的データのビッグデータ化が加速化している昨今において、さらなる研究展開が期待されている分野である。

一般に、生態学や環境学、農学などにおける数理解析は、データ駆動モデル (data-driven model) とプロセスベースモデル (process-based model や mechanistic model) に大別される。ここで、データ駆動型モデルとは、研究対象の応答 (たとえば、生物の在/不在や個体数) とそれを支配する諸条件 (たとえば、気温や標高) の関係性を実測データから発見的に見いだそうとするアプローチであり、機械学習等に基づく高精度な再現と予測が利点である。一方、プロセスベースモデルでは、生物物理モデルに代表されるように、研究対象の挙動について、素過程に関する基礎理論や統計的パターン、経験的知識などに基づいて定式化され、エキスパートシステムや時空間変動モデリング等が代表例として挙げられる。著者の知る限りでは、農業農村工学分野におけるデータ駆動モデルは、ハイドロインフォマティクスに代表される水文学分野を中心に適用されており、次いで、応用生態工学分野や地盤工学分野での研究事例がある。

データ駆動モデルには、統計的な学習法から機械学習を含め多数の手法が存在しており、モデルの構造により再現性が大きく異なることが知られている (Fig. 1)。同図のように、シンプルで分かりやすいモデルは誤差が大きい

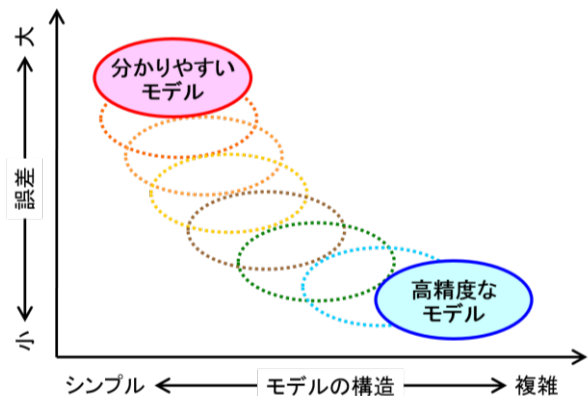


Fig. 1 Accuracy-complexity tradeoff among a variety of models (Ishibuchi, 2007)

¹ 東京農工大学大学院農学研究院 Institute of Agriculture, Tokyo University of Agriculture and Technology, Japan

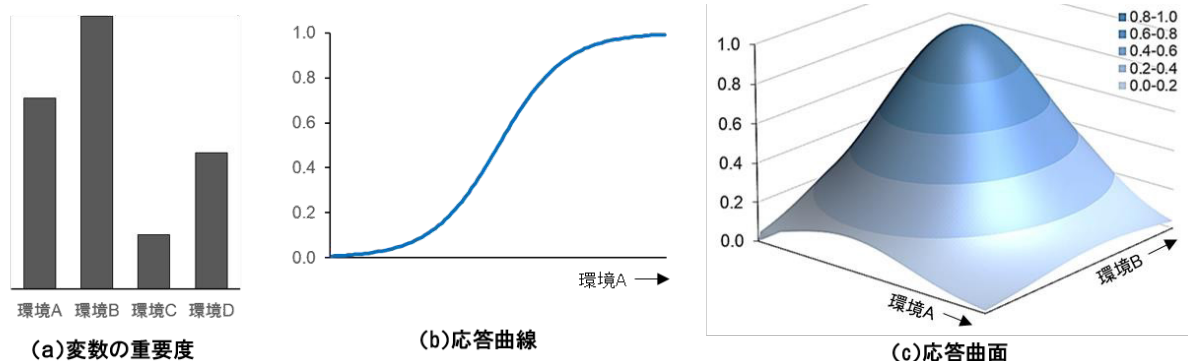


Fig. 2 Conceptual diagram of extracted information from a data-driven model (Fukuda, 2015)

くなる傾向にある一方で、高精度なモデルは構造の複雑さからブラックボックスなモデルとして批判されることがある。それに対し、データ駆動モデルから有用な情報を抽出するための手法が開発されており、変数の重要度や応答曲線等 (Fig. 2) の使用が一般的になりつつある。とくに、応答曲線は生息環境評価の基本となる生息場適性指数 (HSI) として利用できることから、実用上も重要なアプローチである。

3. 魚類生息環境評価への適用例

福田 (2015) で紹介されているように、生態学分野での機械学習の適用は、1990 年代後半に初めてニューラルネットワークが適用されたことから著しい発展を遂げてきた。近年では、フリー統計ソフト R 等の普及により、多様な手法が手軽に利用可能になっており、手法ごとの再現性や抽出可能な生態情報が異なることの報告 (Fukuda *et al.*, 2013) がなされている。その他、機械学習手法の適用上で重要なハイパーパラメータの最適化に進化計算を適用した事例 (Munoz-mas *et al.*, 2018) もあり、手法の高度化が進んでいる。一般に、空間分布モデルは、スナップショット的な空間分布の解析手法であるため、時空間変動 (例えば、移動分散特性) を再現可能な手法の開発は今後の課題の一つである。しかしながら、このような研究には、解析に堪え得るデータが必要であるため、計測やデータ分析等に関する技術革新とともに進展していくものと考えられる。

4. おわりに

本報では、機械学習等を用いたデータ駆動モデルの応用事例を通して、ブラックボックスモデルとして批判されることが多い同モデルから、対象システムに関する情報を抽出する手法について紹介した。今後は、急速に発展する解析技術を随時取り入れつつ、農業農村工学分野における多種多様なビッグデータや高精度・高解像度データへの適用事例を集積し、実地的な維持管理や評価、設計等に資する解析システムの確立が望まれる。

引用文献

- 福田信二 (2015): エコインフォマティクスーデータ駆動型モデルの生態学への応用一. システム/制御/情報, 59(10), 363-368.
- Fukuda, S., De Baets, B., Waegeman, W., Verwaeren, J., Mouton, A.M. (2013): Habitat prediction and knowledge extraction for spawning European grayling (*Thymallus thymallus* L.) using a broad range of species distribution models. Environmental Modelling & Software, 47, 1-6.
- Ishibuchi, H. (2007): Multiobjective genetic fuzzy systems: Review and future research directions. Proc. of 2007 IEEE International Conference on Fuzzy Systems, 913-918.
- Muñoz-Mas, R., Fukuda, S., Pórtoles, J., Martínez-Capel, F. (2018) Revisiting probabilistic neural networks: a comparative study with support vector machines and the microhabitat suitability for the Eastern Iberian chub (*Squalius valentinus*). Ecological Informatics, 43, 24-37.