ミニシンポジウム記録 水産資源データ解析と予測モデル

1. 水産資源学の予測モデル
—モデル選択, 機械学習—北門利英
東京海洋大学1. Prediction in fishery science —model selection
and machine learning—

TOSHIHIDE KITAKADO

Tokyo University of Marine Science and Technology,
Minato-ku, Tokyo 108-8477, Japan

水産資源学において、これまで推測や予測と言えば、主に統計学の枠組が中心であった。最尤法は、尤度を通してパラメータの推測やモデル選択を一貫的に扱うことを可能とした。またベイズ法は、MCMC などによって階層構造などモデルの選択肢や柔軟性を飛躍的に増大させ、さらに不確実性下での意思決定や確率的な言葉による推測結果の提示をも可能としてきた。近年、様々な分野で統計的機械学習を用いた予測モデルが利用され始めているが、水産資源学にもその潮流がみられ始めている。

ところで統計的機械学習によるアプローチとはなんだろうか。統計的機械学習法とは、一定のアルゴリズムを通してコンピュータに学習させ内部的なルールを構築していく方法で、教師付き学習、教師なし学習、強化学習などに分類される。このような学習理論が必要となる状況を整理してみると、①データがスパースである、②データ数が豊富または膨大である、③回帰や分類において想定すべき構造が想定しにくい、などが挙げられる。

①はデータ数と比較して変数の数が多い状況を意味し、例えば回帰分析において候補となりうる説明変数の数が比較的多い場合が例となる。従来、変数選択法としてAIC やステップワイズ検定が、そしてパラメータの推定性能向上法として経験ベイズ法で解釈可能な Ridge 回帰などが利用されていたが、Lasso はこれら2つを同時に扱える手法である。漁海況や漁獲量の予測に、関係しそうな多くの変数を入力する場合などが該当し、利用度も高い。

②は従属変数も多い場合を意味する。通常データ解析を行う場合には1セットのデータしか得られないため、予測の確率的な性能を評価する場合にブートストラップ法などが用いられてきた。リサンプリングというアイデアは、バギングやランダムフォレストというアンサンブル学習理論においても利用される。一つ一つは弱い学習機でも数多くあれば統合することで強くできる、という発想が根底にあり、特に後者は説明変数の次元を縮約してリサンプリングできるため、その性能に加え実用性も高い。

③は因果関係や分類を通して予測を行う際に、パラメトリックな構造を事前に想定できず、それ自身をコンピュータに学習させようというアプローチである。GAM などのノンパラメトリックな方法を想定する状況とも似ているが、結果として得られる構造としては非明示的である。このような考えは、以前からニューラルネットワークやサポートベクターマシンなどで用いられてきた。前者は現在、深層学習という形で生まれ変わって脚光を浴びているが、これは①や②のように情報量の多い場面をも想定した手法となっている。

今回の発表では、予測モデルを切り口に統計的機械学習の概説を行うとともに、水産資源学諸問題へのそれらの適用方法について議論した。今後、CPUE 標準化、ハビタットモデル、空間解析などへの適用が促進されるであろう。

また資源評価モデルにおける典型的な状況として、資源動態モデルの客観的な評価法が必要とされている。そこで予測能力検証型の尺度として、レトロスペクティブなクロスバリデーションを通した予測能力の評価が挙げられる。¹⁾ これにより異なる資源解析モデル間でモデルフリー評価が可能となる。このようなアプローチも予測の別の側面として今後議論が進むべきであろう。

文 献

- 1) Kell L, Kimoto A, Kitakado T. Evaluation of the prediction skill of stock assessment using hindcasting. *Fish. Res.* 2016; **183**: 119–127.