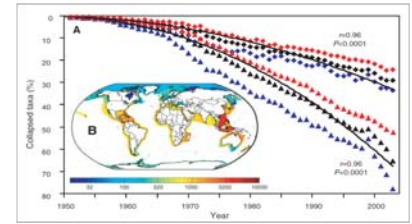


# Functional Boosting---海洋生態学データへの応用

小森 理 統計思考院 特任助教

## 概要

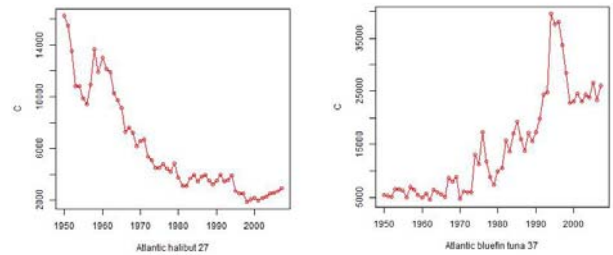
海洋資源評価の問題は、資源の適切な管理運用にとって重要な問題である。Worm et al (2006)の論文では現状の資源管理体制のままでは2048年までに漁業資源の崩壊が示唆されており、海洋資源の評価、管理の見直しの動きが活発になってきている。このような状況の中、近年では個々の種に注目した資源の保全ではなく、世界規模の資源管理が注目を浴びてきている(Thorson et al, 2012; Costello et al, 2012)。この発表ではThorson et al (2012)の論文と同じRAM(Ransom Aldrich Myers)データの漁獲量データ(時系列データ)を用い、資源評価予測(枯渇か否か)をBoostingによって行い、彼らの手法との比較を行った。



Worm et al. (2006)

## RAMデータ

1953年から2008年までのデータ。右図のような漁獲量(catch)データの他に、魚の栄養段階(trophic level), 体長の(Lmax), 漁獲された地域(Fao)の変量もある。資源量かの指標であるRelative Biomassもあり、RB<0.2のときを枯渇と定義する。



## Functional Boosting

2値のクラスラベルをY, m次元の特徴変量をZとし, Zはℓ次元の確率変数Xとm次元の確率関数  $\Xi = \{\Xi(t) : t \in T\}$  に分解できるとする( $Z = \{X, \Xi\}$ ). このときXに対しては通常のstumpを弱学習機として考える。

$$f(x, k, \sigma, \delta) = \sigma \text{sign}(x_k - \delta),$$

ただしk=1, ..., ℓとし, σは±1とする。一方関数データξ(t)に対しては基底関数との内積を考え以下のような弱判別機を用意する。

$$g(\xi, \phi, k, \sigma, \delta) = \sigma \text{sign}\left(\sum_{t \in T} \xi_k(t) \phi(t) - \delta\right).$$

ここで最小化するU-loss functionを以下のように定義する。

$$L_U(F, \mathcal{D}) = \sum_{(x, y) \in \mathcal{D}} U(-yF(x)).$$

但し  $F = \sum \alpha_j f_j$ , 弱判別機とその重みは以下のアルゴリズムで学習する。

$$\begin{cases} F_{\text{new}}(x) = F_{\text{old}}(x) + \alpha_D f_D(x), \\ (\alpha_D, f_D) = \underset{(\alpha, f) \in \mathbb{R} \times \mathcal{F}}{\text{argmin}} L_U(F_{\text{old}} + \alpha f, \mathcal{D}). \end{cases}$$

## 結果

漁獲量の時系列パターンを取り入れたBoostingの手法がAUCで評価して、一番良い結果となった(右図の赤線)。Thorsonの手法はこのような時系列のパターンをうまく判別手法に取り得ておらず、AUCの評価ではかなり悪い結果であることが分かった。今回は基底関数としてB-splineを用いたが、漁獲量の変動にあった基底関数の考察も有用と考えられる。

## 参考文献

- Christopher Costello et al. (2012). Status and Solutions for the World's Unassessed Fisheries, *Science* 338 : 517-520  
 Thorson et al. (2012). Using model-based inference to evaluate global fisheries status from landings, location, and life history data, *Can. J. Fish. Aquat. Sci.* 69: 645-655  
 Worm et al.(2006). Impacts of Biodiversity Loss on Ocean Ecosystem Services, *Science* 314: 787-790

