The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

ベイズ状態空間モデルを用いた定置網漁のための日単位漁獲量予測

幸加木裕也 *1 Kokaki Yuya　堀内優佳 *1 Horiuchi Yuka　俵直弘 *1 Tawara Naohiro　福嶋正義 *2 Hukushima Masayoshi　井戸上彰 *2 Idoue Akira　橋本和夫 *1 Hashimoto Kazuo
小林哲則 *1 Kobayashi Tetsunori　小川哲司 *1 Ogawa Tetsuji

*1早稲田大学 Waseda University　*2KDDI 総合研究所 KDDI Research, Inc.

The daily fish catch forecast is indispensable information that can support fishery workers with their decision-making for efficient operation. Machine learning generally performs well if big data are available. Those data, however, are not always available in developing fish catch forecasting systems. The present study proposes a fish catch forecasting method using a state space model that emulates the process of fixed shore net fishing. Experimental comparisons conducted using actual fish catches and meteorological data demonstrated that the proposed method yielded a significant improvement over general machine learning-based forecasting when only limited amount of data are available.

1. はじめに

定置網漁の仕組みと網内の魚の振る舞いに関する知識を組み込んだ状態空間モデルにより，日ごとの漁獲量を逐次的に予測する漁獲量推定手法を提案する．

日々の漁獲量の正確な予測は，漁業従事者の意思決定，ならびに漁業の効率的なオペレーションに貢献する．例えば，突発的な不漁が予測できれば，その日は定置網の交換を行ったり休漁とするなどの意思決定ができる．逆に突発的な大漁の予測は，漁の開始時間を早めて二度漁に出るための判断材料になり得る．

漁獲量予測の研究は，月や年単位の長期的な漁獲量を予測するものと，1 日ごとの短期的な漁獲量を予測するものに大別できる．長期の漁獲量予測に関しては，これまで盛んに検討がなされてきた [黄 96, 為石 98, 庄野 14] 一方，1 日ごとの漁獲量予測は依然としてチャレンジングな課題であり，研究例は少ない．短期の漁獲量予測として，Leathwick らは，ニュージーランド近海で捕獲された鯛の漁獲量と漁獲確率を水温や塩分濃度から決定木を用いて予測することを試みている [Leathwick 06] 決定木を利用することで，特徴量の各要素の漁獲量予測に対する寄与度が算出可能である．決定木は汎用的な機械学習の手法であり，様々な対象の予測において優れた精度が得られることで知られているが，予測器を学習するために大規模なデータを必要とする．現状，漁獲量予測においては大規模かつ整備されたデータが常に用意できるとは限らないため，機械学習アプローチによらない新たな予測手法が求められている．

そこで，本研究では，定置網漁の漁獲量予測において，少ない学習データ量でも頑健に動作する予測器として，定置網内の漁獲過程を表現した状態空間モデルを提案する．状態空間モデルでは，観測データの背景にある生成過程を潜在変数を用いてモデル化し，観測データを最もよく説明できる生成過程をデータから推論することで予測を行う．このとき，予測対象の系に関する知見を用いて状態空間モデルを定式化することで，少量データでも頑健に動作する予測器を構築することができる．このように対象とする系に関する様々な知見を用いて状態空間モデルを構築し解析を行う研究として，生物の個体群の推定 [Iijima 13] や日次食料需要の分析 [松田 15] などがあるが，漁獲量推定にこのようなアプローチを適用した例は著者の知る限り存在しない．

本稿では，既存の決定木を用いた予測方式と，提案する状態空間モデルに基づく予測方式を東松島沖に設置された定置網で漁獲されたシロサケの漁獲量推定に適用することで，各手法における学習データ量と予測性能の関係性を調査する．

以降の本稿の構成は以下の通りである．2. では，汎用的な機械学習の手法である決定木を用いた漁獲量予測モデルについて述べた後，定置網内の漁獲過程を表現した状態空間モデルを提案する．3. では 2 つの予測方式について，学習データ量を変化させた際の予測性能を調査する．最後に 4. において全体の総括を述べる．

2. 定置網漁の漁獲量予測モデル

本節では，定置網漁の漁獲量予測方式について述べる．2.1 では決定木を用いた予測方式について述べ，2.2 では，提案する状態空間モデルに基づく予測方式について述べる．

2.1 決定木による漁獲量予測

過去の漁獲量および気象・海洋情報から当日の漁獲量を予測するモデルを gradient boosting decision tree (GBDT) を用いて構築する．ここでは，漁獲量は概ね前日と同様であり，そこからの変動は当日の気象・海洋情報 (表 2) により決まると仮定する．そこで，過去 N 日間 (実験では $N = 7$ とした) の漁獲量と気象・海洋情報を入力として，前日からの漁獲量の増減値を出力する GBDT を構築する．なお，予備実験として，GBDT の出力を前日からの差分ではなく漁獲量そのものとした場合も評価を行っており，GBDT の出力を前日からの差分とするモデルのほうが高い性能を達成することがわかっている．

2.2 定置網漁の漁獲過程を表現した状態空間モデル

定置網内における魚の確率的な振る舞いを状態空間モデルとして定式化する．

2.2.1 定置網漁の仕組みと定式化

定置網の構造を図 1 に示す．定置網は，網への魚の流入を促す垣網や運動場網などの部分，流入してきた魚を留めておく箱網 (第 1 箱網)，箱網内に留まった魚を確実に取得するための金庫網 (第 2 箱網) の 3 つの部分から成る．運動場網から箱

連絡先: kokaki@pcl.cs.waseda.ac.jp

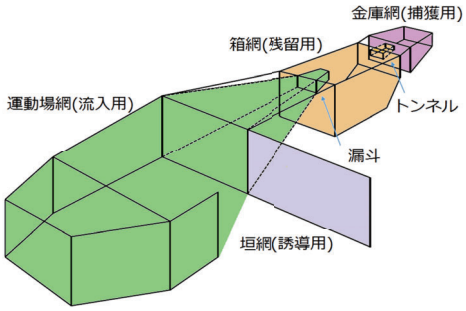

図 1: 定置網の構造

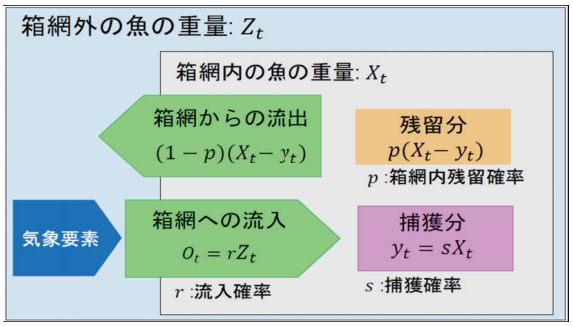

図 2: 箱網における魚の入出力関係

網へは移動がし易く，箱網から運動場網への流出は難しい構造となっている．そのため，箱網内には魚が滞留し，その一部が奥の金庫網に流入する．漁では，漁船から金庫網を巻き上げるので，金庫網に流入した魚のほとんどは捕獲に成功することを意味する．

図 2 は，定置網漁における漁獲の過程，特に箱網における魚の入出力についての概念図である．ここで，Z_t は時刻 t に網周辺に存在する魚の総重量，X_t は時刻 t に網内に存在する魚の総重量とする．このとき，網の入出力関係を表 1 に示す．

状態空間モデル

このような漁獲の過程に基づいて状態空間モデルを定義する．状態方程式 (1) は，箱網内の魚の総量 X_t を状態としたときの，魚の網への流入および流出の過程を表す．また，観測方程式 (2) は，金庫網から魚が捕獲される過程を表す．

$$X_{t+1} = \mathcal{N}(p(X_t - y_t), k(X_t - y_t)p(1-p)) + \mathcal{N}(rZ_t, kZ_t r(1-r)) \quad (1)$$

$$y_{t+1} = \mathcal{N}(sX_{t+1}, kX_{t+1}s(1-s)) \quad (2)$$

なお，$\mathcal{N}(\mu, \sigma)$ は，平均 μ，分散 σ をパラメタとするガウス分布である．式 (1) の第 1 項は，前の時刻からの網内に残留する魚の総量を表し，第 2 項は網外からの流入分を表す．このうち観測されるのは漁獲量 y_t のみであり，網内の魚の量 X_t，網周辺の魚の量 Z_t，網への残留確率 p，網への流入確率 r，漁獲確率 r は，推定すべきモデルパラメータとする．いま，各項が確率分布で表現されているが，これは表 1 に示した仮定によって流入・流出の過程がポアソン過程で記述されるためである．また，網内への流入・流出が起こる回数は十分に大きいと考えられるため，ポアソン分布はガウス分布によって近似できるとする．なお k は，同一行動を選択する魚 (魚群) の総重量 (魚一個体の平均重量 × 魚群の構成数) の平均値を表す係数である．ここでは，個体の平均重量 5 [kg] と仮定し，捕獲されるかどうかの試行は個体ごとに行われるとすると，$k = 5$ [kg] となる．

表 1: 箱網における魚の入出力の詳細

漁獲	網内の魚はそれぞれ確率 s (漁獲確率) で金庫網に移動する (捕獲される) と仮定する．
残留	箱網内で捕獲されなかった魚はそれぞれ確率 p (残留確率) で網内に残留すると仮定する．
流出	箱網内で捕獲されなかった魚はそれぞれ確率 $1-p$ (残留の余事象) で網内に残留すると仮定する．
流入	網周辺の魚の群れは確率 r (流入確率) で網内に流入すると仮定する．

実装

t 日目における網周辺から網内への流入量 $o_t = r \cdot Z_t$ は，漁獲量予測日の前日までの H 日間のうち，実際に漁に出た d 日の平均漁獲量 M_t に比例すると仮定すると，以下の式が成り立つ．

$$o_t = r \cdot Z_t = q_t \cdot M_t \quad (3)$$

$$M_t = \frac{1}{d} \sum_{i=1, y_{t-i} \neq 0}^{H} y_{t-i} \quad (4)$$

M_t は観測データから計算できる量で，実験では $H = 7$ とした．また，q_t は，網内への魚の流入度合いを表すパラメータであり，後述する方法で推定する．この値を使うことで，式 (1) の第 2 項は以下のように書きなおすことができる．なお，r はその定義から 1 より十分小さな値であると考えられるため, 実装では $1 - r \simeq 1$ と近似した．

$$\mathcal{N}(rZ_t, kZ_t r(1-r)) = \mathcal{N}(q_t M_t, k' M_t q_t) \quad (5)$$

ここで，魚群の構成数を G 匹と仮定すると，k' は，$k' = k \times G$ [kg] となる．実験では，$G = 100$ とし，$k' = 5 \times 100 = 500$ [kg] とした．また，式 (1) の第 1 項における $X_t - y_t$ は X_t の推定値によっては 0 以下になるため，ガウス分布の分散が定義できなくなる．この問題を回避するため，$X_t - y_t$ を $(1-s) \cdot X_t$ で置き換えた．この定式化の下で網内の魚の量 X_t，網内への流入度 q_t をデータから逐次推定する．なお，残留確率 p，漁獲確率 s は日時に依らず一定とする．

網内への流入度 q_t の算出

風が強く吹けば潮の流れが速くなる．鮭は潮の流れに抗うように動き回るため，その結果として網に入り易くなる，と言われている．これは，漁師の経験からよく知られていることである．そこで，網への流入度 q_t は気象の影響を受けると仮定して，気象情報を用いて最適な q_t を推定する枠組みを導入する．まず簡単なモデルとして，q_t は気象情報に陽に依存せず q_{t-1} からのランダムウォークとして決定されるモデルを考える．この場合，気象情報の影響はランダムウォークとして表現される．このとき，q_t は，分散 σ^2 を用いて以下のように書ける．

$$q_t \sim \mathcal{N}(q_{t-1}, \sigma^2) \quad (6)$$

次に，気象情報を陽に用いて q_t を推定するモデルとして，過去数日の漁獲量と気象条件から q_t を GBDT により推定するモデルを次式で定義する．

$$q_t = \mathrm{GBDT}(Y_{t-D:t-1}, W_{t-D:t-1}) \quad (7)$$

表 2: 使用した気象・海洋情報の詳細

属性	詳細
気温 (3)	平均気温，最高気温，最低気温
湿度 (3)	平均蒸気圧，平均湿度，最小相対湿度
降水量 (2)	降水量の合計，10 分間降水量の最大値
気圧 (3)	平均現地気圧，平均海面気圧，最低海面気圧
降雪 (2)	最深積雪，降雪量合計
風速 (3)	平均風速，最大風速，最大瞬間風速
風向 (3)	最多風向，最大風速風向，最大瞬間風速風向
その他 (1)	日照時間

ここで，$Y_{t-D:t-1}$ は過去 D 日分の漁獲量を要素に持つベクトル，$W_{t-D:t-1}$ は過去 D 日分の気象・海洋情報を成分に持つベクトルである．本研究では，$D = 7$ とした．

2.2.2 HMC 法を用いた漁獲量の逐次一期先予測

Hamilton Monte Carlo (HMC) 法 [Duane 87] による未知パラメータの推定と日単位の逐次漁獲量予測を同時に行う手法を提案する．提案モデルの未知パラメータは X_t，q_t，σ^2 である．ただし，q_t を GBDT により推定する場合は HMC 法により推定する未知パラメータは X_t のみである．これら未知パラメータは，D 日前から前日までの漁獲量 $y_{t-D:t-1}$ を用いて，HMC 法によりベイズ推定した．HMC 法で得られた未知パラメータのサンプリング値 $X^{*(n)}_{t-:t-1}$，$q^{*(n)}_{t-D:t-1}$，$\sigma^{2*(n)}$ を用いると，前日までの漁獲量に対する当日の漁獲量の事後確率 $p(y_t|y_{t-D:t-1})$ は次式により解析的に評価できる．

$$\begin{aligned} p(y_t|y_{t-D:t-1}) &= \int p(y_t|X_{t-D:t-1}, q_{t-D:t-1}, \sigma^{2*(n)}) \\ &\quad \cdot \mathrm{d}p(X_{t-D:t-1}, q_{t-D:t-1}, \sigma^{2*(n)}|y_{t-D:t-1}) \\ &\simeq \frac{1}{N}\sum_{n=1}^{N} p(y_t|X^{*(n)}_{t-D:t-1}, q^{*(n)}_{t-D:t-1}, \sigma^{2*(n)}) \end{aligned} \tag{8}$$

ただし，N は HMC 法により生成した未知パラメタの仮説数である．また，最終項の漁獲量 y_t の尤度は，状態空間モデルの式 (1-2) と式 (6) について全ての未知パラメタをサンプリング値に置き換えることで解析的に評価できる．Alg. 1 に，HMC 法を用いて漁獲量の推定値とその信頼度を逐次求める手順を示す．

3. 漁獲量予測実験

決定木を用いた予測方式と，提案する状態空間モデルに基づく予測方式について，学習データ量と予測性能の関係性を調査した．

3.1 実験データ

宮城県の東松島沖で定置網漁を営む漁業従事者より，定置網漁による鮭の漁獲量を得た．2011 年，2012 年，2014 年，2015 年，2016 年，2017 年の各々に対して 10 月 1 日〜11 月 30 日の漁獲量データを取得した．気象情報として，定置網の設置場所に近い宮城県の石巻観測台にて観測された情報を用いた．表 2 に示した 20 種類を気象庁のホームページ [JMA] より取得した．

3.2 実験方法

比較する漁獲量予測モデルは以下の 2 種類である．提案手法を **CFM**(Catch Forecasting Model) と呼ぶ．

表 3: 学習データ量と予測精度の関係（RMSE[kg]）

学習データ量	GBDT	CFM
1 年分	1518.2	**1218.7**
2 年分	1230.6	**1198.6**
3 年分	1163.2	**1161.0**
4 年分	1137.8	**1119.1**
5 年分	1128.7	**1125.1**

- **GBDT**: GBDT を用いた漁獲量予測モデル
- **CFM**: 定置網の漁獲過程を表現した状態空間モデル．q_t は式 (7) に従って与えられるとする．

これら 2 つの予測モデルについて，学習に用いるデータ量の予測性能への影響を調査した．

予測性能は，root mean squared error (RMSE) により評価した．RMSE は以下の式 (9) により算出される．

$$\mathrm{RMSE} = \sqrt{\sum_{t=1}^{T} \frac{(y_t - \hat{y}_t)^2}{T}} \tag{9}$$

ここで，y_t は観測された漁獲量であり，$\hat{y}_t$ は漁獲量の推定値である．

3.3 実験条件

GBDT に基づく漁獲量予測モデル (**GBDT**) は，過去 7 日間の漁獲量と表 2 に示した 20 種類の気象情報計 146 次元を入力として，前日からの漁獲量の差分を予測する．

一方，網への流入度 q_t を GBDT を用いて推定するモデル (**CFM**) においては，過去 7 日間の漁獲量および気象情報計 146 次元を用いて, q_t を推定する GBDT を学習させる必要がある．q_t の教師データは，q_t が式 (6) により与えられるとした場合に，1 シーズンの漁獲量データを通して推定される q_t の系列とした．

提案モデルのパラメータ p および s は，ともに 0.7 とした．この値は，p および s に対して 5 通りの値 0.1，0.3，0.5，0.7，0.9 の組み合わせを考え，RMSE を評価することで実験的に決定した．定置網内の魚の存在比に関する調査 [秋山 99] によれば，箱網と金庫網に入っている魚のうち，金庫網内に存在する魚の量の割合は 0.7 程度という知見が得られている．つまり，今回の設定値 $s = 0.7$ は，現実に即した値となっている．

3.4 実験結果

GBDT と **CFM** について，学習データ量と予測精度の関係を表 3 に示した．このとき，1 年分の学習データを利用する場合，予測年と学習年の組み合わせは 30 通り存在する．表に記した予測精度は，学習に用いた年と評価に用いた年の全ての組み合わせの平均値である．表 3 より，**CFM** の学習データ量減少に伴う予測精度の低下は，**GBDT** に比べると穏やかであり，学習データを 5 年分利用したときは予測精度の RMSE の差が 3 kg 程度であったのに対し，学習データを 1 年分しか利用しなかったときはその差が 300 kg 程度に増加している．

図 3 は 1 年分の学習データを用いて 2015 年の日次漁獲量を予測した結果である．学習データの選び方は 5 通りあるため，それら 5 通りの予測値の平均値を予測値とした．この結果から，**GBDT** の予測結果のピークを与える日は正解値と一致しておらず，また，シーズン終盤に実際の漁獲量よりも 1000 kg 以上大きな予測値が得られたことがわかる．一方，**CFM** の

Algorithm 1 HMC 法を用いたベイズ推定による逐次一期先予測の手順

逐次一期先予測アルゴリズム ※ $y_{:t-1}$ を漁獲量の時系列とする．
繰り返し: $n = 1, 2, \cdots, N$:(N は仮説数である．)
パラメタ推定: HMC 法を用いて提案手法の未知パラメタ $\boldsymbol{\theta}^{(n)}_{:t-1}$ を漁獲量の時系列からベイズ推定する．
仮説ごとの予測: $p_r(y^{(n)}_t|\boldsymbol{\theta}^{(n)}_{:t-1})$ の最尤値を n 番目の仮説における漁獲量の予測値 $y^{(n)}_t$ とする．
出力: N 個の仮説のうち $warmup$ に当たる時期の予測値を棄却したもの $\{y^{(n)}_t\}^N_{n=warmup}$ について，
それらの中央値を漁獲量の予測値とする．
また $\{y^{(n)}_t\}^N_{n=warmup}$ において 15 パーセンタイルと 85 パーセンタイルを求め，予測値の信頼度とする．

繰り返し: $t = 8$ からシーズンの終わりまで逐次一期先予測を繰り返す．
If $t < D$ **then:** $y_{1:t-1}$ の下で逐次一期先予測アルゴリズムを適用．
Else: $y_{t-D:t-1}$ の下で逐次一期先予測アルゴリズムを適用．

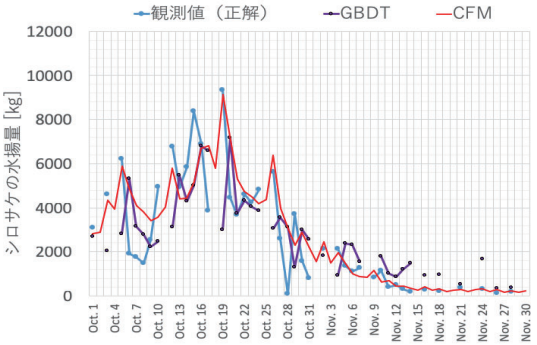

図 3: 学習データ量 1 年分のときの予測結果（2015 年）

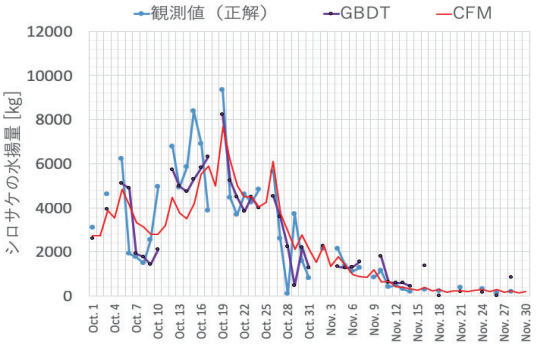

図 4: 学習データ量 5 年分のときの予測結果（2015 年）

予測結果のピークを与える日は正解値と一致しており，シーズン終盤においても **GBDT** に比べ誤差が少なくなっていることがわかる．このように，1 年分の学習データがしか利用できないなどの場合においては，提案する定置網の漁獲過程を表現した状態空間モデルが特に効果的であることがわかる．

図 4 は 5 年分の学習データを用いて 2015 年の日次漁獲量を予測した結果である．この結果から，**GBDT** でも漁獲量のピークを正確に予測できたことがわかる．またシーズン終盤において，その時期に相応しくない極端に大きな予測値を出力することも少なくなっている．このように既存の決定木による予測手法は，十分な量の学習データがあれば，高い精度で漁獲量を推定できることがわかる．

4. まとめ

状態空間モデルにより定置網の漁獲過程を表現した定置網漁専用の予測モデルと決定木により構築した汎用の予測モデルについて，学習データ量が予測精度に与える影響を調査した．実験結果より，学習データ量が十分なときは 2 つの予測手法は同程度の性能を示すが，学習データが少量であるときは，定置網漁に関する事前知識をモデルに取り込むことで，予測の信頼性を向上できることがわかった．

謝辞

本研究開発は総務省 SCOPE（受付番号 172302010）の委託を受けたものです．

参考文献

[黄 96] 黄康錫: ニューラルネットワークを用いた米神定置漁場のマアジ漁獲量予測, 水産海洋研究, Vol.60, No.2, pp.136-142 (1996)

[為石 98] 為石日出生, 小松輝久, 青木一郎, 杉本 隆成: ニューラルネットワークを利用した東北海域のカツオ漁獲量予測, 水産海洋研究, Vol.62, No.4, pp.327-333 (1998)

[庄野 14] 庄野宏, 堀江昌弘, 井上あゆみ, 東剛志: 機械学習に基づく鹿児島近海に来遊するクロマグロ幼魚の漁獲量予測, 計量生物学, Vol.35, No.1, pp.1-15 (2014)

[Leathwick 06] Leathwick,J.R., Elith,J., Francis,M.P., Hastie,T. ,Taylor,P.: Variation in demersal fish species richness in the oceans surrounding New Zealand: an Analysis using boosted regression trees, Marine Ecology Progress Series, Vol.321, pp.267-281 (2016)

[Iijima 13] Iijima,H., Nagaike,T., Honda,T.: Estimation of deer population dynamics using a Bayesian state-space model with multiple abundance indices, The Journal of Wildlife Management, Vol.77, No.5, pp.1038-1047 (2013)

[松田 15] 松田敏信: 状態空間モデルによる日次食料需要の分析, 農業生産技術管理学会誌, Vol.21, No.4, pp.101-111 (2015)

[Duane 87] Duane,S., Kennedy,A.D., Pendleton,B.j., Roweth,D.: Hybrid Monte Carlo, Physics Letters B, Vol.195, No.2, pp.216-222 (1987)

[JMA] ttp://www.jma.go.jp/jma/indexe.html

[秋山 99] 秋山清二: 定置網各部の魚群量割合, 日本水産学会誌, Vol.65, No.1, pp.60-65 (1999)