

AIとデータサイエンス： どのように社会実装に結びつけるか

松原 仁(東京大学AIセンター／
公立はこだて未来大学)

2021年10月9日

松原 仁

- * 1986年 東大大学院工学系研究科情報工学専攻博士課程修了 工学博士
- * 1986年 通産省工業技術院電子技術総合研究所(現産業技術総合研究所)入所
- * 2000年公立はこだて未来大学教授
- * 2014－2016年 人工知能学会会長
- * 2020年 東京大学AIセンター教授&公立はこだて未来大学特任教授
- * 専門:人工知能全般、ゲーム情報学, AIと創造性, 社会実装など

人工知能

- * Artificial Intelligence AI
- * 明確な定義はない(知能を定義することが人工知能の目標?)
- * 人工知能の意味は時代によって変わっている
- * 人間のような知性を持った人工物(コンピュータ、ロボット)を作ること为目标とする(工学的な目的)
- * コンピュータを題材にして知能について研究する(科学的な目的)
- * 個人的には鉄腕アトムを作りたい

人工知能の歴史

- * 1940年代 コンピュータができる
- * 1950年代—1960年代 人工知能スタート
- * 1956年 AIという名称がつく 一回目のブーム
- * 1960年代—1970年代 一回目の冬の時代
- * 1980年代—1990年代 二回目のブーム エキスパートシステム
- * 1990年代—2000年代 二回目の冬の時代
- * 2010年代— 三回目のブーム ディープラーニング
- * 2020年代— そろそろブームが落ち着きつつある？

AIとデータ

- * AIの2回目のブームは人間(の専門家)の知識(ヒューリスティックス)をルールなどの形で明示的にコンピュータに載せようとした
- * 知識を明示的に取り出すことがむずかしかった(人間は意識していない知識をたくさん持っている)
- * AIの3回目のブームは人間(の専門家)が行なっていることの結果(データそのもの)から(暗黙的でもいいので)知識を取り出すことにそれなりに成功している

思考ゲームの場合の数

	場合の数	コンピュータが勝った年
* チェッカー	10の 30 乗	1994年
* オセロ	10の 60 乗	1997年
* チェス	10の 120 乗	1997年
* 将棋	10の 220 乗	2017年
* 囲碁	10の 360 乗	2017年

* 数学的に解けた(必勝法がわかった)のはチェッカーまで

Deep Blue(1)

- * IBMが1989年から開発を開始
- * 1990年世界チャンピオンのカスパロフと
対戦 2戦2敗
- * 1996年再度カスパロフと対戦(フィラデルフィアに見に行った)
6戦 1勝3敗2分け

Deep Blue(2)

- * 1997年 ニューヨーク
6戦 2勝1敗3引き分け
1秒間に2億手 3分で14手先読み
スーパーコンピュータ
+ チェス専用コンピュータ512台





チェスプログラムとデータ

- * チェスは機械学習を本格的に使わずに世界チャンピオンに勝った
- * 過去の主要な棋譜データを記憶して用いていた
- * 最近は(将棋や囲碁と同様に)大量の棋譜データから学習している

ボナンザメソッド

- * ボナンザ以前のコンピュータ将棋
- * 評価関数は人間が手作業で作成
- * 探索は前向き枝刈り
- * ボナンザ(2005年) 保木邦仁(電通大)
- * 評価関数を機械学習で作成
- * 玉を含む3駒関係 KPP(King Piece Piece)
要素数が億に及ぶ
- * 探索は全数探索が基本

最適制御理論(1)

- * 最大(小)化問題として力学系の制御方針を推論する
- * ラグランジュ形式の解析力学
- * パルス整形による化学反応制御
- * 最小燃料のロケット弾道
- * 工場の消費電力
- * 池の魚に与える餌

最適制御理論 (2)

- * $J = \int l(x, u, t) dt$

- * t : 時間に関する数 (離散も可)

- * $x(t)$: 系の状態

- * u : 制御変数

- * t を手数, $x(t)$ を minimax 探索の指し手で発展していく局面, u を特徴ベクトルとみなし, 最適制御理論に基づいた将棋プログラムの機械学習をおこなう

第二回電王戦

- * 2013年3月23日 阿部4段対習甦 ×
- * 3月30日 佐藤4段対ponanza ○
- * 4月6日 船江5段対ツツカナ ○
- * 4月13日 塚田9段対Puella α △
- * 4月20日 三浦8段対GPS将棋 ○

コンピュータがプロ棋士に3勝1敗1引き分けで勝利！！
(現役男性プロ棋士に初勝利)

終了宣言

- * 2015年10月11日 情報処理学会将棋プロジェクト終了宣言。トッププロ棋士に勝つコンピュータ将棋を開発するというプロジェクトの目的を事実上クリアできたので終了する。マスコミでかなり取り上げられる。

将棋プログラムとデータ

- * ボナンザまでは序盤の定跡データのみ
- * ボナンザは人間の棋譜約10万局のデータから機械学習 KPPからKPPT (King Piece PieceTurn)
- * その後NNUE (Efficiently Updatable Neural-Network-based evaluation functionsの逆順)系が主流に
- * 最近は DL(Deep Learning)系に移りつつある
- * 深層強化学習 (Deep Reinforcement Learning)

アルファ碁

- * 2016年1月 アルファ碁 発表 Nature
- * ディープラーニング＋強化学習＋モンテカルロ木探索
- * 強化学習 (正と負の)報酬に基づく機械学習 古い歴史あり
- * ディープラーニング 人間の打った3000万局面をデータとして学習させた

アルファ碁(2)

- * 最初の発表のときはプロ2段に勝った
- * 2016年3月に韓国のイ・セドル(元世界一で当時のトップの一人)と対戦して4勝1敗 世界中が驚いた
- * 2017年5月 中国のカ・ケツ(世界一)と対戦して3勝0敗
- * これでアルファ碁は引退

アルファ碁ゼロ

- * 棋譜は入力しない(人間の囲碁の知識を用いない)
 - * 囲碁のルールだけ教える
 - * 強化学習を用いる
-
- * アルファ碁よりも強くなった
 - * 画期的と評価された

囲碁プログラムとデータ

- * 初期は序盤の定石データのみ
- * モンテカルロ囲碁は統計処理
- * アルファ碁は深層学習 (3000万局面の棋譜データ)
- * 深層強化学習

コンピュータ麻雀

- * ネット麻雀の普及によって進展
- * とつげき東北「科学する麻雀」(講談社現代新書)2004
- * これがデータを用いた麻雀研究の始まり
- * 「デジタル派」の台頭
- * かなりの格言(たとえば「早いリーチは愚形待ち」)は統計的には誤り
- * 爆打 水上直紀
- * Super Phoenix Microsoft Asia
- * 人間のトップレベルに並ぶ
- * とつげき東北「新 科学する麻雀」(ホビージャパン)2021

コンピュータに俳句を作らせたい

- * 北海道大学が中心(川村秀憲教授)
- * 「AI一茶」プロジェクト
- * 数十万の俳句をコンピュータに覚えさせる
- * ディープラーニングで学習させる
- * 言葉を並べて17文字にして、季語1つ、切れ字1つ以下のものを選ぶ
- * コンピュータは意味はわかっていない

人間対AI俳句対決

(2018年7月13日北海道大学)

金葎屍の後へ置く小花(人類の1句目) 8点
花蜜柑剥く子の道の地平まで(AIの1句目) 6.5点
馬蛤貝の波につまづき潮に巻く(人類の2句目) 7点
撒くといふ言葉正して花見ゆる(AIの2句目) 5点
許しがたい臭いを放屁虫(人類の3句目) 5.75点
無人とは毛深くなりし狸かな(AIの3句目) 6.75点
仮名の裏がえりをそむ子ら梅雨晴間(人類の4句目) 6.25点
山肌に梟のこげ透きとほる(AIの4句目) 5点
ホルン吹く放課後の大夕焼かな(人類の5句目) 8点
かなしみの片手ひらいて渡り鳥(AIの5句目) 8.5点

2019年3月9日対決

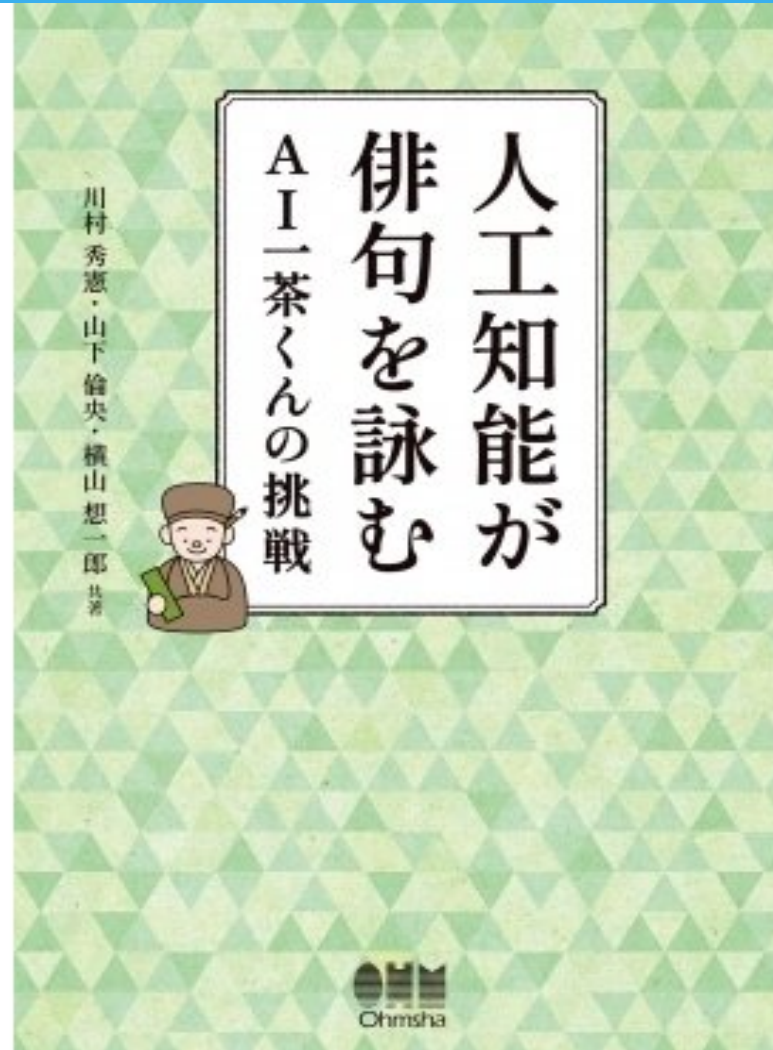
※掲句・作者一覧

試合記録										
第5試合		第4試合		第3試合		第2試合		第1試合		試合
朧		落の臺		菜の花		雛		冴返る		兼題
紅	白	紅	白	紅	白	紅	白	紅	白	チーム
踊り場に鍵取り落す朧かな	どこまでも産当たりたる朧かな	落の臺散らしてみたる会釈かな	開店のチラシ来たらし落の臺	菜の花の速達で来るラブレター	菜の花や世界となりてひとつづつ	雨音の途切れずしづか古雛	戦争がだんだん昏れて雛祭	朝シヤンのやうな顔して冴返る	ダイヤルの電話の黄ばみ冴返る	掲句

2019年3月9日対決結果

※掲句・作者一覧

試合記録											
第5試合		第4試合		第3試合		第2試合		第1試合		試合	
朧		菜の臺		菜の花		雛		牙返る		兼題	
紅	白	紅	白	紅	白	紅	白	紅	白	チーム	
踊り場に鍵取り落とす朧かな		とくまでも崖当たりたる朧かな		菜の花の速達で来るラブレター		雨音の途切れすしづか古雛		朝シャンのやうな顔して牙返る		ダイヤルの電話の黄ばみ牙返る	
朱子		一茶904		ハル		麻衣		一茶2395		たんぽぽ	
一茶225		幹生		一茶13		一茶508		掲句		作者	



AIの社会実装とデータ

- * データは現場にある
- * 北海道はデータの宝庫
- * 農業
- * 交通
- * 漁業
- * 観光

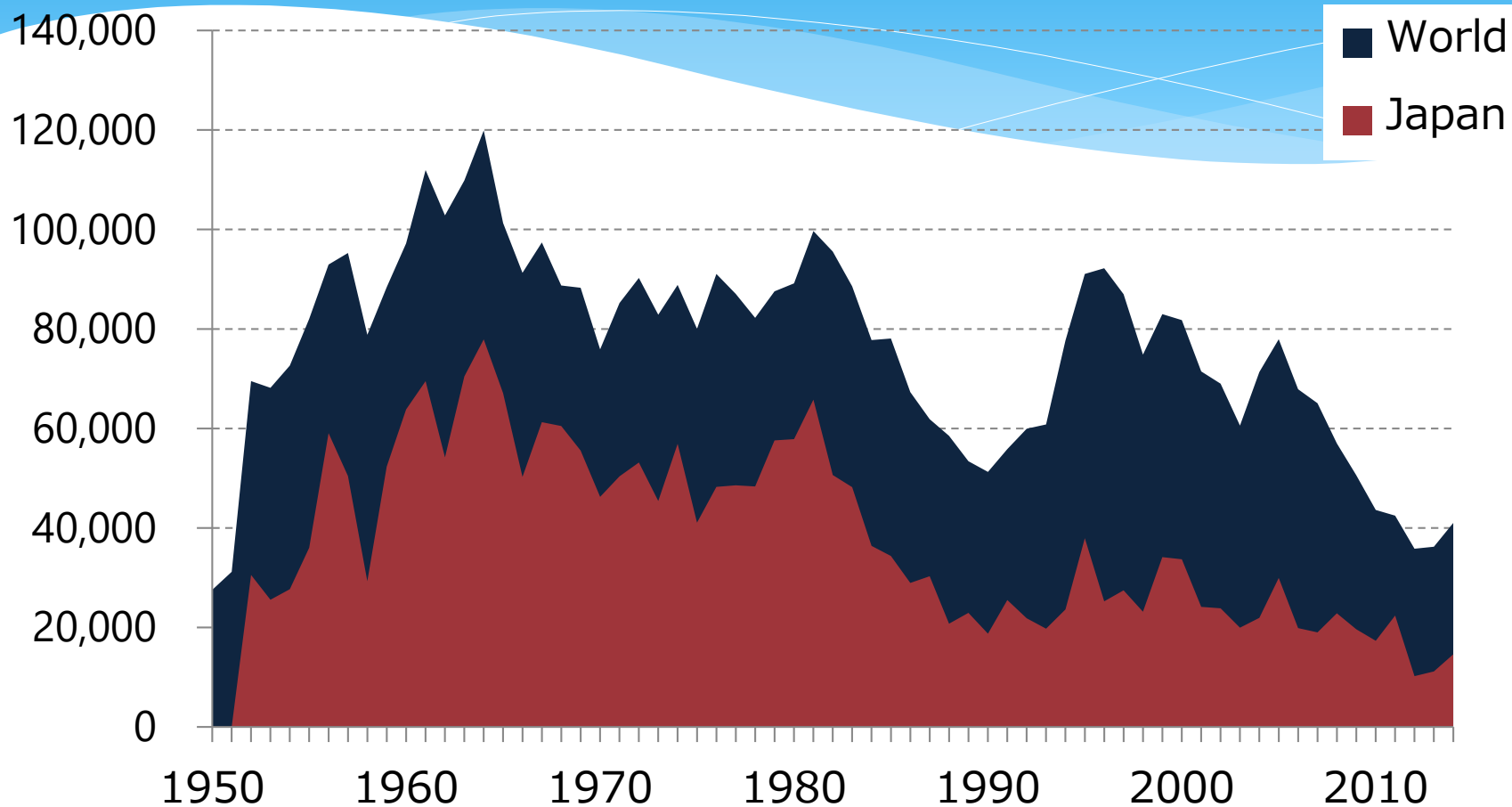
定置網漁業



- 定置網漁業の特徴
 - ✓ 受動的な漁法
 - * 環境にやさしい漁法
 - ✓ 漁獲選択性はない
 - * 色々な魚が獲れる
- めじまぐろの資源管理
 - ✓ 6ブロックでの資源管理
 - * 太平洋北部ブロックは上限超過
 - * 昨年6月に10日間の禁漁を実施
 - * 大きな社会問題になっている



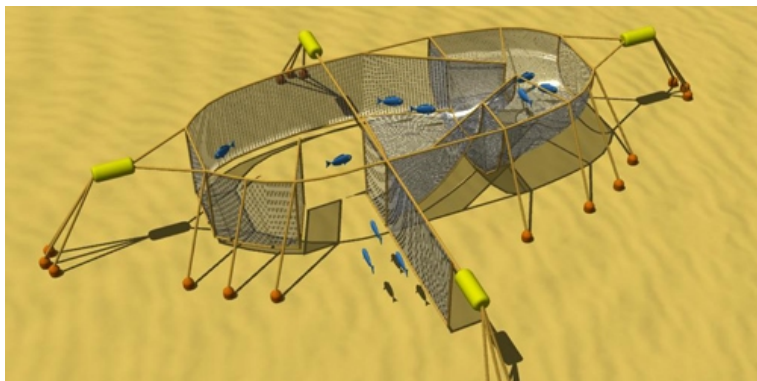
まぐろの漁獲量の推移



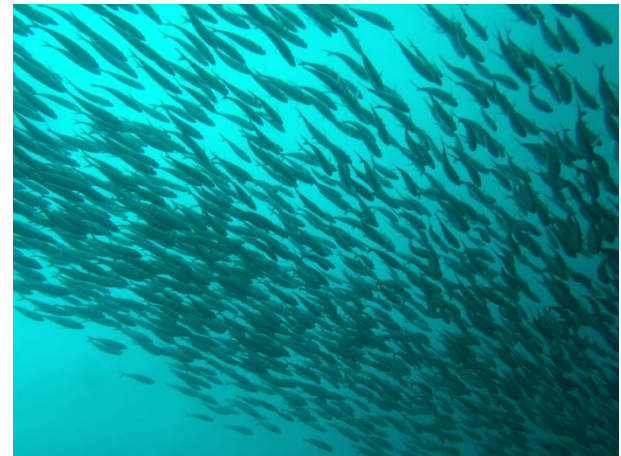
定置網の経営体数



	全国	北海道	函館
大型定置網	431	34 (7.9%)	13
さけ定置網	821	821 (100%)	17
小型定置網	2,867	552 (19.3 %)	7
合計	4,119	1,407 (34.2 %)	37



定置網魚群探知機



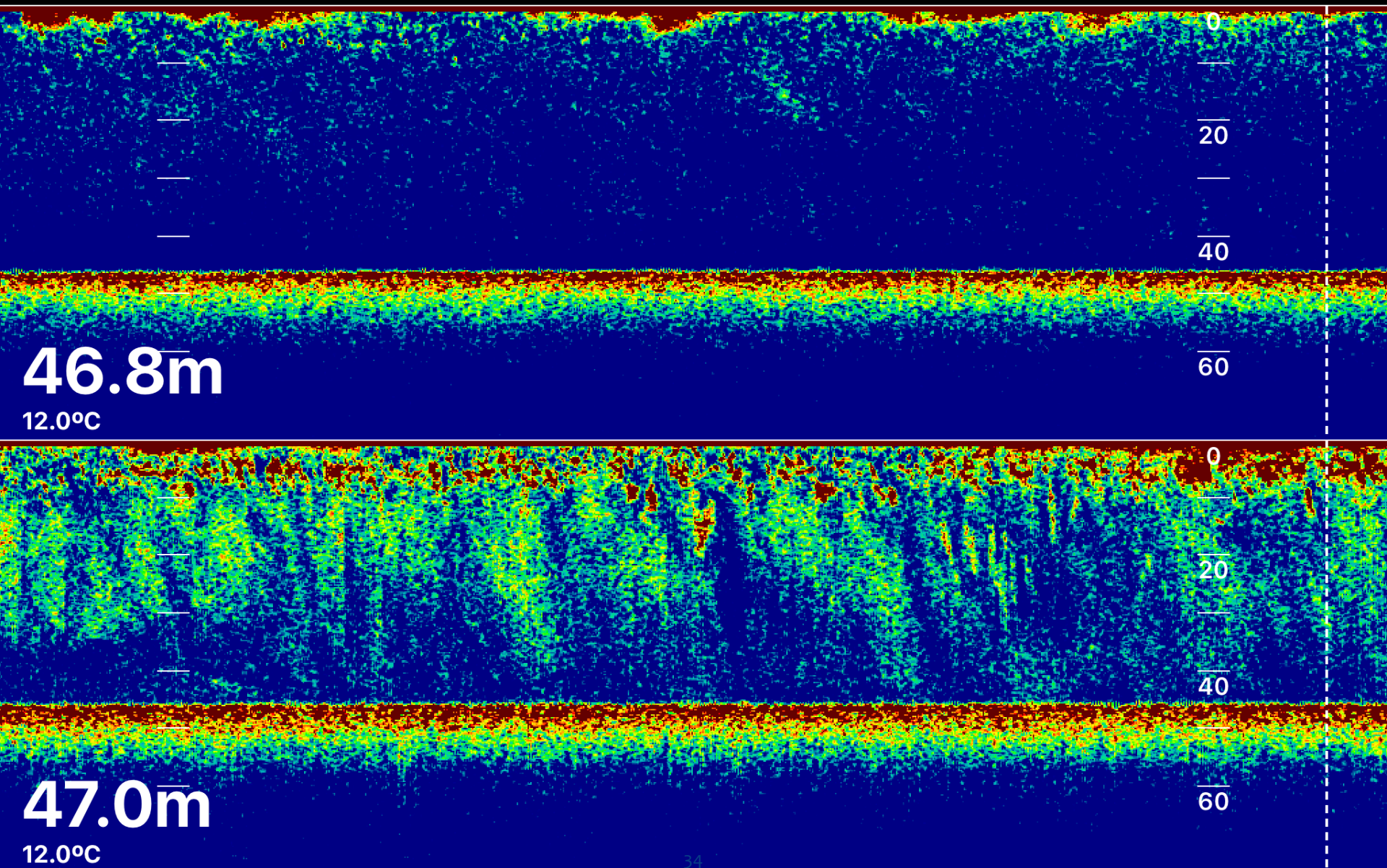
KODEN

0.07kt(NNW) 13.1°C

O:120 G:100 V:12.53



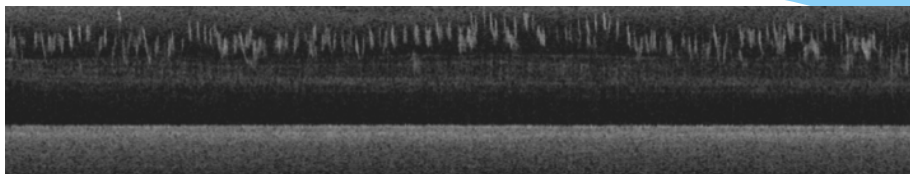
2016/10/28 04:02



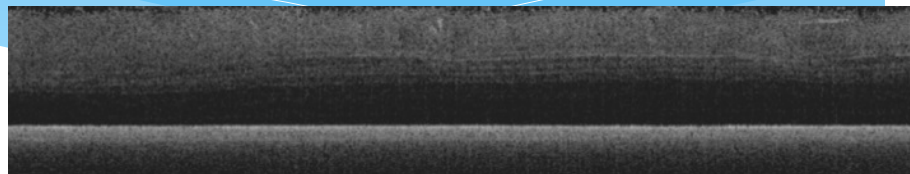
魚種毎のグレースケール画像



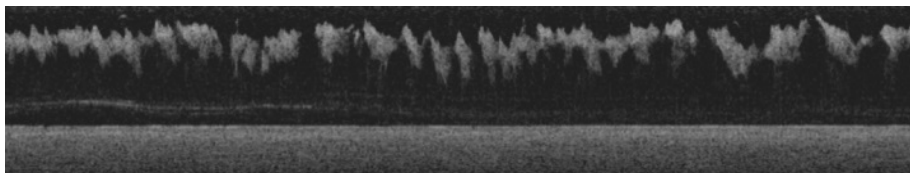
めじまぐろ 2,684kg (10月3日)



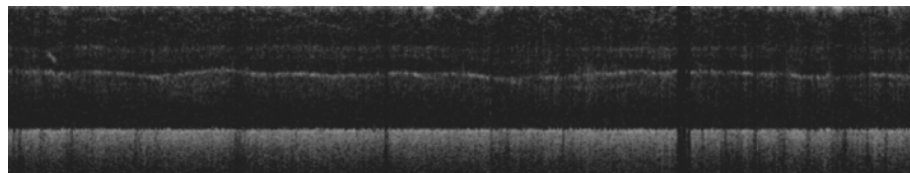
さけ 1,560kg (11月5日)



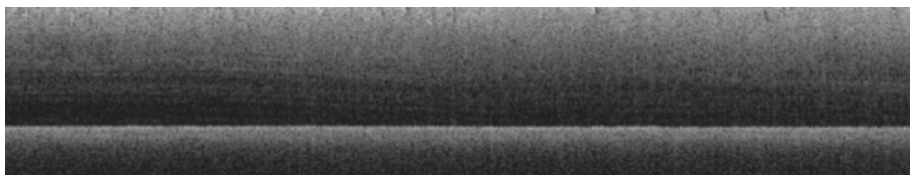
ぶり 2,835kg (10月6日)



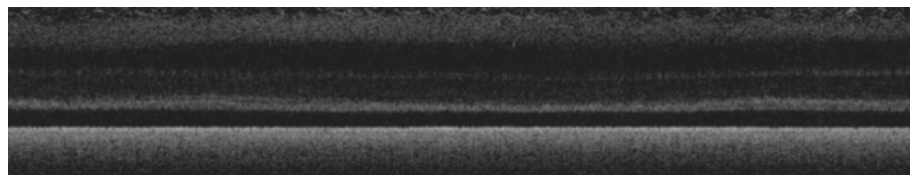
するめいか 2,990kg (11月11日)



かたくちいわし 4,045kg (11月1日)



漁網 133kg (10月22日)



資源保護の効果



● 2015年10月の漁獲データに適用

- ✓ 出漁日数 29日間
- ✓ 漁獲量 96,170kg (5,911kg)
- ✓ 仮想休漁 2日間

	保護量	漁獲量
めじまぐろ (5,911kg)	3,365kg	2,546kg
その他 (90,259kg)	2,725kg	87,534kg

マリンAI

- * 定置網の魚群探知機データを定期的に収集
- * 毎日の水揚げのデータを収集(教師データ)
- * それらのデータをディープラーニングで学習
- * 網の中にどの魚がいるかを判定できるようになる

観光とデータとAI

- * 観光情報学会 2003年設立(2009－2014年 会長)
- * 情報を切り口として観光を捉えてその振興を図る
- * データが大量に存在する
- * 膨大な情報から傾向を見出す
- * 膨大な情報から適切なものを取り出す
- * 観光を振興するために必要な情報は何かを探す
- * 観光に共通する要素とその観光地特有の要素を切り分ける

まとめの代わり

- * AIの社会実装にとってデータが必要不可欠
- * データを集める方法は進歩している(センサー、インターネット、デジタル化など)
- * データを扱う技術がますます重要になる
- * データサイエンスを誰にいつどう教えるかが重要