

1. 取り組んでいる研究
2. 研究の予定
3. 進捗状況

研究の進捗

海野幸也 (Yukiya Unno)

富山県立大学 工学部
情報システム工学科

May 19, 2025

ホタルイカの身投げ量を予測

ホタルイカが取れる条件として様々な条件が挙げられている。

- シーズン：3～5 月（例年）
- 風向き：南風
- 気温：日中暖かい日
- 月齢：月明りのない「新月」
- 水色：潮にニゴリがない状況
- 潮回り：大潮前後

過去の口コミから、その日にどれだけのホタルイカが身投げされたかのデータを取得し、その身投げ量と、その日の天気、気温、月齢、風などからどのような条件で身投げが起きやすいかを分析する。その情報から、どの日が身投げ量が多いかの予報が載っているサイトを作成する。

背景

富山県のホタルイカの身投げは、産卵行動や環境要因と密接に関連する現象であり、その量の分析と予測は、生態学的理解の深化や気候変動の指標、さらに地域観光や漁業資源管理における重要な意義を持っている。

目的

- 1 過去のホタルイカ身投げデータ（量、天気、気温、月齢、風など）を収集
- 2 身投げが起きやすい条件の分析
- 3 身投げ量の予報を提供するウェブサイトの作成

データの取得方法

過去の身投げ量データの取得: ホタルイカ掲示板というサイトにどれだけ身投げがあったかの口コミがあり、その口コミをスクレイピングで取得し、身投げ量、日付、場所を csv にまとめる。

過去の天気、風、気温などのデータの取得: 気象庁の公式サイトに、場所を指定したら、過去の天気などが詳しく載っている。そこからスクレイピングでデータを取得する。

伏木（富山県） 2017年4月3日 （1時間ごとの値）

時	気圧(hPa)		降水量(mm)	気温(°C)	露点温度(°C)	蒸気圧(hPa)	湿度(%)	風向・風速(m/s)		日照時間(h)	全天日射量(MJ/m²)	雪(cm)		天気	雲量	視程(km)
	現地	海面						風速	風向			降雪	積雪			
1	1015.9	1017.5	--	3.4	-1.3	5.5	71	2.6	南西			--	--	Ⓐ		20.0
2	1015.6	1017.2	--	3.4	-2.3	5.1	66	3.1	南西			--	--	Ⓐ		20.0
3	1015.8	1017.4	--	3.6	-2.8	5.0	63	3.0	南西			--	--	Ⓐ		20.0
4	1016.0	1017.6	--	3.2	-2.7	5.0	65	2.3	南西			--	--	Ⓐ		20.0
5	1016.0	1017.6	--	3.1	-2.8	5.0	65	2.7	南西			--	--	Ⓐ		20.0
6	1016.8	1018.4	--	3.2	-2.7	5.0	65	3.0	南西	0.2		--	--	Ⓐ		20.0
7	1017.0	1018.6	--	5.2	-2.1	5.2	59	2.3	南南西	1.0		--	--	Ⓐ		20.0
8	1017.1	1018.7	--	7.0	-0.7	5.8	58	2.4	南西	1.0		--	--	Ⓐ		20.0
9	1016.6	1018.2	--	11.3	0.1	6.2	46	4.7	南西	1.0		--	--	Ⓐ		17.2
10	1016.0	1017.6	--	11.9	0.7	6.4	46	5.8	南南西	0.3		--	--	Ⓐ		20.0
11	1015.5	1017.1	--	12.2	0.9	6.5	46	7.3	西南西	0.3		--	--	⊙		20.0
12	1014.3	1015.9	--	13.0	2.6	7.3	49	6.7	南南西	0.0		--	--	⊙		20.0

天気予報データの取得: 気象庁から提供されている天気予報 API（完全無料で使える API）を使用してデータを取得する（予定）。

月齢データの取得: 月齢は、西暦と日付から計算できるので、自分で API を作る予定。

分析方法（検討中）

回帰分析

- 取れ高と環境要因（気温、月齢など）の関係を回帰分析で調べる。
- 単回帰分析や重回帰分析を使用し、予測モデルを構築。
- 回帰係数を基に取れ高を予測する。

機械学習

- 過去のデータを用いて機械学習モデルを構築。

過去の身投げ量データの取得

- 1 ホタルイカ掲示板というサイトの口コミからスクレイピングで口コミ内容を取得
- 2 その口コミ内容から、ホタルイカの身投げ量を 5 段階に分類

ホタルイカ掲示板というサイトの口コミからスクレイピングした。コメント数は全部で 14000 件ほどある。2015 年から 2025 年までのコメントで、約 900 日分。

岩瀬 はな 投稿日: 2025年04月20日 01:35

現時点ホタルイカ15杯ほど取れました。

無題 さくらきちひろき 投稿日: 2025年04月20日 01:29

魚津漁港奥
人なし 濁りなし いかなし ふぐあり

朝日病院裏 昌幸 投稿日: 2025年04月20日 00:50

濁りあり波強しイカ無し人少し

1. 取り組んでいる研究
2. 研究の予定
3. 進捗状況

最初は、口コミ内容を Sentence BERT を使用し分類したが、精度が悪かった。

```
日付：2025年03月30日 01:57  
コメント：波高し  
潮満ちはじめ  
イカは無し  
分類結果：なし  
  
日付：2025年03月30日 01:52  
コメント：10分で50匹ぐらいのペースですね  
分類結果：なし  
  
日付：2025年03月30日 01:25  
コメント：岩瀬  
波多し  
人多し  
イカゼロ  
駄目だな今回は。  
分類結果：なし  
  
日付：2025年03月30日 01:21  
コメント：いません。人だけが大量です。  
分類結果：非常に多い  
  
日付：2025年03月30日 01:14  
コメント：岩瀬浜  
諦めて車内でするめ食ってる  
分類結果：不明
```

長い文章や、ホタルイカ以外がことを書いてある場合に精度が悪くなった。

その後、Sentence BERT を使用したファインチューニングを行い精度を上げようとしても、精度が上がらなかった。調べてみると、ファインチューニングには1 ラベルにつき 1000 件以上の学習データが必要らしい。今回は 5 段階のラベルがあるため、学習用のデータは 5000 件以上のデータが必要になる。

1. 取り組んでいる研究
2. 研究の予定
3. 進捗状況

AI (google の geminiAI) の API を使用し分類を行なってもらったら精度がかなり上がったので、これを使用することにした。

```
prompt = """
# 富山のホタルイカの漁き委員会分類タスク

## 指示
以下のテキストを分析し、富山のホタルイカの漁き委員会（身投げ量）を最も適切なカテゴリに分類してください。

## 分類カテゴリと定義
- 「なし」：ホタルイカが全く観察されていない
  - キーワード例：全くいない、見られない、0匹、皆無、成果なし
- 「少ない」：わずかな数のホタルイカが観察された
  - キーワード例：わずか、数匹、ちらほら、ボツボツ、1〜10匹程度
- 「普通」：平均的な数のホタルイカが観察された
  - キーワード例：そこそこ、まあまあ、10〜50匹程度
- 「多い」：かなりの数のホタルイカが観察された
  - キーワード例：たくさん、いっぱい、50〜100匹程度、多く見られる
- 「非常に多い」：異例に大量のホタルイカが観察された
  - キーワード例：すごい数、大量、総満ち、300匹以上、1キロ以上
- 「時期不明」：コメントが置かれた時期のリアルタイムな情報ではない
  - 「昨日」「先週」など過去の情報
  - 「見逃げ旅」など過去の漁師に関する情報
- 「不明」：漁き委員会について明確な情報がない、または判断できない

## ルール
1. まず、テキストがホタルイカの漁き委員会に言及しているか確認する
  - 言及がない場合は「不明」と分類
2. 次に、情報がリアルタイムかどうかを確認する
  - リアルタイムでない場合は「時期不明」と分類
  - テキスト内に漁獲情報がない場合は、その情報はリアルタイムな情報であるとする
3. リアルタイムな漁き委員会情報がある場合のみ、量に基づいて5段階（なし、少ない、普通、多い、非常に多い）で分類する
4. 「人の多さ」「量の多さ」「漁獲量」などだけの情報では「不明」とする。ただし、イカに関する記述がある場合はそれを優先して分類する。

## 分析対象テキスト
{Text}

## 出力形式
分類結果を「なし」「少ない」「普通」「多い」「非常に多い」「時期不明」「不明」のいずれか一つだけ出力してください。説明文や追加コメントは含めなくてください。
"""
```

このようなプロンプトを投げた。

1. 取り組んでいる研究
2. 研究の予定
3. 進捗状況

```
{
  "date": "2025/04/27 02:38",
  "comment": "ボツボツしか取れてません...\n自己中な人が多く、居駐も多く、無理やり入って来る人も多い",
  "surge_level": "少ない"
},
{
  "date": "2025/04/27 02:27",
  "comment": "イカたくさんでます。予想通りわいた1時間からわいてます",
  "surge_level": "多い"
},
{
  "date": "2025/04/27 02:22",
  "comment": "大量にとってる方もチラホラ",
  "surge_level": "多い"
}
```

約 14000 コメントを分類し、日付とコメント内容と分類結果を json 形式で保存した。

分類結果として、

- 不明: 6820 件
- なし: 3654 件
- 少ない: 2180 件
- 普通: 555 件
- 多い: 1001 件
- 非常に多い: 470 件

という結果が返ってきた。研究で使用するデータは、不明以外のの計 7860 データを使用する。

1. 取り組んでいる研究
2. 研究の予定
3. 進捗状況

分類したデータと月齢データを結合し、月齢との関係のみを簡単に分析してみた

ロコミを分類したデータの json に、自作した月齢 API を使用し、月齢データを結合した。

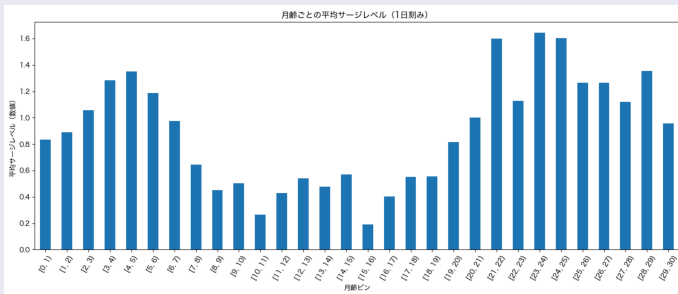
```
{
  "date": "2025年05月05日 01:02",
  "comment": "ホタルイカいません！\n人はぼつり。\\n今年はどうもイカちゃん来ないですかねー？",
  "surge_level": "なし",
  "moon_age": 8.59
},
{
  "date": "2025年05月05日 04:03",
  "comment": "2時半ごろに奇跡の1匹のみ。次は月末、入替、朝日、糸魚川移動になりそうです。もう、明るくなってきました。",
  "surge_level": "少ない",
  "moon_age": 7.59
},
{
  "date": "2025年05月05日 03:09",
  "comment": "ホタルイカなし\\nヤリイカ1",
  "surge_level": "なし",
  "moon_age": 7.59
},
}
```

進めたところ（続き）

11/12

1. 取り組んでいる研究
2. 研究の予定
3. 進捗状況

なし:0, 少ない:1, 普通:2, 多い:3, 非常に多い:4 とそれぞれの口コミの分類結果を数値化し、月齢ごとの口コミの分類結果の平均を計算した。その結果をグラフにした。



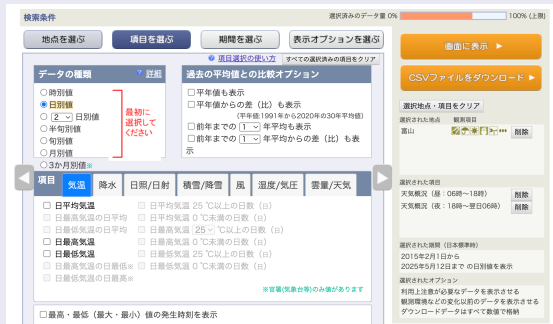
満月は月齢 14.8、新月は月齢 0.0 である。

月齢が満月付近では口コミの分類結果が低くなる傾向があることがわかった。逆に月齢が新月付近の時に、口コミの分類結果が高くなる傾向があることがわかった。

1. 取り組んでいる研究
2. 研究の予定
3. 進捗状況

天気、風、気温、降水量などのデータの取得

調べていると気象庁の公式サイトに、過去の天気、風、気温、降水量などのデータがあり、指定した場所、期間を指定すると、過去の天気などを CSV 形式でダウンロードできることがわかった。



検索条件

選択済みのデータ量 0% 100% (上限)

地点を選ぶ 項目を選ぶ 期間を選ぶ 表示オプションを選ぶ

項目選択の使い方 すべての選択済みの項目をクリア

データの種類の 詳細

☐ 時別値

☒ 日別値

☐ [2~] 日別値

☐ 半旬別値

☐ 旬別値

☐ 月別値

☐ 3か月別値

過去の前平均値との比較オプション

☐ 平均値も表示

☐ 平均値からの差 (比) も表示

☐ 前年までの (1~) 年平均も表示

☐ 前年までの (1~) 年平均からの差 (比) も表示

項目

気象 降水 日照/日射 積雪/降雪 風 湿度/気圧 雲量/天気

☐ 日平均気温

☐ 日平均気温 25℃以上の日数 (日)

☐ 日最高気温の日平均

☐ 日平均気温 0℃未満の日数 (日)

☐ 日最低気温の日平均

☐ 日最高気温 (25℃以上) の日数 (日)

☐ 日最高気温

☐ 日最高気温 0℃未満の日数 (日)

☐ 日最低気温

☐ 日最低気温 25℃以上の日数 (日)

☐ 日最高気温の日最低

☐ 日最低気温 0℃未満の日数 (日)

☐ 日最低気温の日最高

※気象(気象台等)のみ値があります

☐ 最高・最低 (最大・最小) 値の発生時刻を表示

画面を表示

CSVファイルをダウンロード

選択された項目をクリア

選択された地点 観測項目

山梨

選択された項目

天気概況 (昼: 06時~18時)

天気概況 (夜: 18時~翌06時)

選択された期間 (日本標準時)

2015年2月1日から

2025年5月12日までの 日別値を表示

選択されたオプション

利用上注意が必要なデータを表示させる

観測環境などの変化以前のデータを表示させる

ダウンロードデータはすべて数値で格納

今後はそのデータを json に変更して使用し、口コミの分類結果と結合し、分析などを行う予定。