

# 天気・気温・湿度・月齢から機械学習した昆虫採集数推定手法

今林 仁広

Yoshimasa IMABAYASHI

## 1 はじめに

カブトムシやクワガタムシなどの昆虫は日本人にとってとても馴染み昆虫であり夏の風物詩とも揶揄されている。夏場になると山や林にカブトムシやクワガタムシを目的とした昆虫採集に出かける一般人や家族連れも多く見受けられる。しかし山や林に行くことで、これらの昆虫を採集することは困難である。専門的な知識を有する有識者と、そうではない一般人とでは成果に大きな隔たりが存在することが昆虫採集の現状である。

有識者は昆虫採集の際に一般人とは異なり、採集時期や採集日の状況を深く考慮する。それは夏場といっても細かな時期によって採集可能な昆虫の種類は異なるためである<sup>1)</sup>。また、昆虫採集の成果はその日の状態によって大きく異なる。昆虫採集を行う際に考慮すべき因子として、天候、気温、湿度、月齢が挙げられる。そこで本研究では、一般人が昆虫採集において求める種類ごとに関するの個体発生状況を、天気・気温・湿度・月齢といった環境因子を用いて明らかにする。これにより昆虫採集に関する知識を持たない一般人の採集成果向上を図る。実際の昆虫採集でも使用可能なかどうか決定木解析とベイズ推定により明らかにする。そして自然環境への情報処理技術の応用の有用性を示す。

## 2 昆虫採集数推定手法

### 2.1 昆虫採集数推定手法の概要

昆虫採集数推定手法は、昆虫採集を行う際に考慮すべき因子として挙げられる、天気・気温・湿度・月齢を機械学習して、採集日当日にどの種類がどの程度採集可能なのか推定するものである。

### 2.2 昆虫採集数推定手法の構成

昆虫採集数推定手法の構成として、採集日当日に採集可能なのか判断する採集可否推定、採集可能ならばどの程度採集可能なのか推定する採集数推定の2つから構成される。

採集可否推定は採集数データと環境因子からベイズ推定により採集可能かどうか判断する。採集日のデータ  $d$  が与えられたとき、それがどのクラス（採集可能日もしくは採集不可能日）  $c$  に属する可能性が高いのか考えたとき、条件付き確率  $P(c|d)$  はベイズの定理より、

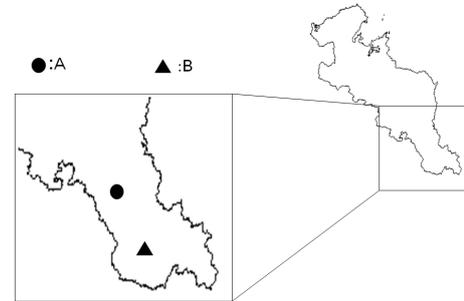


Fig. 1 採集地

(1) 式で計算する事ができる。

$$P(c|d) = P(c)P(d|c)/P(d) \quad (1)$$

(1) 式により、右辺を各クラスつまり、採集可能日/採集不可能日それぞれについて計算し、採集日がどちらに属するのか求める。

採集数推定は採集数のデータと環境因子から決定木を作成し、どのような環境の時にどの程度の採集量が期待できるのか推定する。このとき決定木を作成するアルゴリズムとして CART(classification and regression tree) アルゴリズムを用いた。採集目標種に応じて決定木を作成し、採集量を推定する。

## 3 データ収集と分類器および決定木の作成

### 3.1 データ収集実験と採集環境

昆虫採集数は著者が採集した数をデータ数として扱う。採集時間は原則として日暮れ以降の2時間以内としている。これは長時間に渡り昆虫採集を行って採集量が極端に増加してしまう事を防ぐ為である。昆虫採集で用いる道具は懐中電灯と1.5mの網とする。一般人の昆虫採集を想定しているために特殊な道具もしくは工具は一切用いないとする。

考慮すべき要素である気温・湿度は TANITA 製の CPW-125 を用いて計測する。計測手法として採集地の任意の定点にフィールドワーク直前に設置し採集後に示される数値を記録する。天気は晴・曇・雨の3つのパターンに分類した。

採集地域は Fig. 1 で示す、A 地点：京都府京田辺市甘南備山周辺ならびに、B 地点：京都府木津川市神童寺周辺の2地点とした。

そして本研究において、2015年度と2016年度の6月から9月にフィールドワークを実施し、計2115匹の個体を採集することができた。収集した環境因子のデータをTable 1に示す。これらのデータを用いて各推定を行う。

環境因子	観測値
天気	晴・曇・雨
気温	16℃～33℃
湿度	56%～99%
月齢	0～15

### 3.2 分類器および決定木の作成

データ収集実験で得られたデータよりシステムを構築するための分類器および決定木の作成を行う。採集可否推定の分類器として2015年度の採集データを用いる。特徴量は天気・気温・湿度・月齢とし、採集可能か採集不可能のグループに分類を行う。採集数推定の決定木として、2015年度と2016年度の採集データを用いる。特徴量は上記と同様に天気・気温・湿度・月齢とし、採集数を0匹・1～4匹・5～9匹・10匹以上の4つのグループに分類を行う。

## 4 昆虫採集数推定手法の精度検証

### 4.1 精度検証の概要

採集可否推定分類器における精度検証として、2015年度に得ることができた採集データを訓練データとし、2016年度の採集データが採集可能日なのか採集不可能日なのか分類を行う。採集数推定における推定精度検証として、2015年度と2016年度の採集データより決定木を作成し、1個のデータ(1日/86日)をテストデータとし、残りを訓練データとして一個抜き交差検証を行う。

### 4.2 精度検証の結果

#### 採集可否推定分類器における精度検証結果

分類結果をTable 2に示す。5種類全てにおいて分類率が80%以上となった。しかし分類率が最も高いコクワガタの100%から最も低いノコギリクワガタの81.2%のように分類率に差があった。このような原因としてコクワガタが採集された日数が、全採集日86日のうち84日と採集率が高いことが考えられる。事前確率により求められたコクワガタが採集される環境に全て正しくトレーニングデータが分類されたために、分類率が100%となった。一方で分類率が最も低いノコギリクワガタに関しては、採集されたが採集数が少ない日と採集されなかった日の分類が正しく行えなかったことが考えられる。ミヤマクワガタもノコギリクワガタと同様の理由に

より分類率が84.3%で推移したものと考えられる。

Table 2 ベイズ推定による分類結果

種類	分類率 [%]
カブトムシ	93.5
コクワガタ	100.0
ノコギリクワガタ	81.2
ミヤマクワガタ	84.3
ヒラタクワガタ	90.9

#### 採集可否推定分類器における精度検証結果

Table 3に各種類において、全86日の採集日のうち推定値と実際の採集値つまり満足度が一致した割合を示す。各種類において51.1%から75.5%の精度で推移しており、全種類において50%以上の精度で採集前にユーザの満足度を推定していることが分かった。推定値よりも実際の採集量が多い場合においても誤分類としてしまうために、ノコギリクワガタの精度が50%代を推移したと考えられる。本論文では86日分のデータを用いたが、複数のユーザが本手法を用いることでデータ量が増え、精度も向上すると考えられる。

Table 3 交差検証による各種類の精度

種類	分類精度 [%]
カブトムシ	60.8
コクワガタ	62.7
ノコギリクワガタ	51.1
ミヤマクワガタ	75.5
ヒラタクワガタ	67.4

## 5 結論

本研究は昆虫採集の知識を持たない一般人が有識者のように安定した採集量が得られないといった課題に対し、採集量を推定する昆虫採集数推定機能を提案した。この機能を用いることにより昆虫採集の知識を持たない一般人でも、採集成果を推定することが可能である。

2015年度と2016年度の2ヶ年、計86日分のデータを用いて機能の精度検証を行った。検証結果より、ユーザ志向における本手法は採集対象種全てにおいて50%以上の精度を持つことが分かった。さらに採集日当日に、採集可能か採集不可能かといった指標に対しては、採集対象種全てにおいて80%以上の精度を持つことも分かった。このことから昆虫採集に関して知識を持たない一般人においても効率的な昆虫採集が可能であると言える。

## 参考文献

- 1) 馬場金太郎, 平嶋義宏森本 康司, "馬場金太郎 1991 昆虫採集学, "昆虫採集学, 九州大学出版会, pp.812, 1991