

確率勾配ブースティングを用いた耕作放棄の要因分析・拡大予測モデルの開発

Developing a diagnostic and predicting model of cultivated land abandonment with gradient boosting decision tree

左右田篤^{*1}
SOUDA Atsushi松井孝典^{*1}
MATSUI Takanori福井健一^{*2}
FUKUI Ken'ichi町村尚^{*1}
MACHIMURA Takashi大阪大学大学院工学研究科^{*1}
Graduate School of Engineering, Osaka University大阪大学産業科学研究所^{*2}
Institute of Scientific and Industrial Research, Osaka University

In Japan, cultivated land abandonment has been increasing in recent years, upsetting natural symbiotic systems in farmlands. Therefore, it is required to identify the factors of abandonment and forecast future abandonment. In this study, we developed a process to produce the diagnostic and predicting model of abandonment with high accuracy by gradient boosting decision tree. The result shows that the model is highly accurate for predicting cultivated land abandonment rate after five years.

1. 序論

日本では耕作放棄地が年々と増加しており、生態系や食料自給率などに悪影響を及ぼしている。したがって、耕作放棄の圧力が高い地域を特定するとともに、促進要因の緩和策を提案する意思決定支援システムの開発が求められる。これに対して[松井 2014]では機械学習を用いて地域別の耕作放棄の要因分析と予測を同時に行えるモデルを開発したが、予測の精度の改善とより詳細な要因の特定の必要性が指摘されている。そこで本研究では、経営耕地の変化と耕作放棄の動態を個別に高精度に将来予測し、その要因分析が可能なモデル構築プロセスの開発を目的とする。

2. 耕作放棄拡大予測モデルの構築プロセス

2.1 対象地の選定

農業が盛んな地域で耕作放棄地面積率が全国平均に近い水準である、福井県、石川県、富山県の北陸 3 県を対象地に、旧市区町村を基礎として 526 のサンプルを選定した。

2.2 分析用データの収集

耕作放棄拡大予測モデルの説明変数と応答変数を抽出するため、農林水産省が公開している農林業センサス[農林水産省 2017]のうち販売農家の営農情報と、国土数値情報[国土交通省 2017]の 5 次メッシュスケール (250m²)での標高と傾斜、市街化区域などの土地計画区分のデータを使用した。

2.3 予測モデルの構造の設計

式 (1) により i 年の耕作放棄地面積率 $AFLR_i$ を予測した。

$$AFLR_i = \frac{AFL_i}{TFL_i + AFL_i} \quad (1)$$

ここで AFL_i および TFL_i は i 年の耕作放棄地面積 [ha]、経営耕地面積 [ha] を表す。本研究では最終的に $AFLR_i$ を予測するために、 AFL_i を予測する耕作放棄予測モデル A、 TFL_i を予測する農地縮小予測モデル T を、式 (2) と式 (3) のようにそれぞれ構築した。 j 年は i 年の 5 年前を表す。

連絡先: 左右田篤, 大阪大学大学院工学研究科環境・エネルギー工学専攻, 565-0871, 大阪府吹田市山田丘 2-1, atsushi.souda@ge.see.eng.osaka-u.ac.jp

$$AFL_i = f_A(x_j, \nabla x_{j,k}) \quad (2)$$

$$TFL_i = f_T(x_j, \nabla x_{j,k}) \quad (3)$$

ここで x_j は j 年の説明変数、 $\nabla x_{j,k}$ は k 年から j 年での説明変数の変化率を表す。 k 年は j 年の 5 年前を表す。説明変数は 2 つのモデル共通で、農林業センサスから収集した販売農家の i 年、 j 年、 k 年の共通項目、ならびに国土数値情報から収集した標高と傾斜の平均、標準偏差、最大値、最小値と、土地計画区分の面積割合を採用した。これらの説明変数は、表 1 に示すように「労働力要因」、「経済要因」、「農業経営要因」、「地理要因」、「開発圧力要因」の大きく分けて 5 つのカテゴリに分類される。欠損値処理と VIF 基準によるデータの加工を行うことで、県別ダミーを含めて計 164 の説明変数が分析用として採用された。

2.4 確率勾配ブースティングの適用

本研究では、識別性能が高く、かつ要因を特定できる必要があるため、式 (2) の f_A 、式 (3) の f_T に弱学習器に回帰木を用いた確率勾配ブースティングを適用した。10×10 の入れ子交差検証によるグリッドサーチでベストパラメータを決定した。実装には、R ver.3.32 上で xgboost と caret のライブラリを用いた。

2.5 汎化性能の評価

本研究では耕作放棄予測モデル A と農地縮小予測モデル T を構築するため、式 (2) と式 (3) について、 $k = 2000$ 、 $j = 2005$ 、 $i = 2010$ でトレーニングを行い、 $k = 2005$ 、 $j = 2010$ 、 $i = 2015$ をテストデータとしてモデルの汎化性能を評価した。

表 1 説明変数のカテゴリの定義と変数の数

カテゴリ	説明	変数
経済要因	農業収入や農地などの所有財産など農家の経済状況に関する要因	15
労働力要因	所有労働力及びそれらの年齢や性別など農家が所有する労働職に関する要因	54
農業経営要因	作付け作物の種類やその栽培方法、農地の貸し借りなど農地の経営に関する要因	85
地理要因	標高、傾斜角度など地域の地理に関する要因	4
開発圧力要因	市街化地域などの開発圧力に関する要因	5

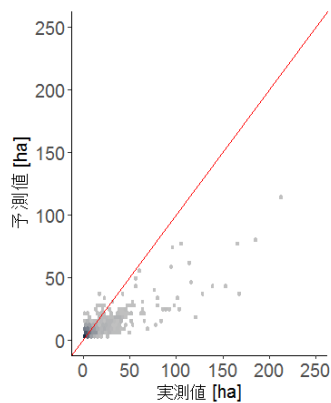


図1 AFL₂₀₁₅の予測値と実測値の比較

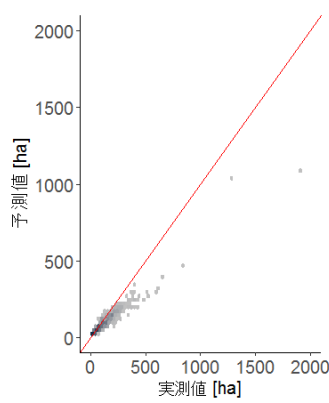


図2 TFL₂₀₁₅の予測値と実測値の比較

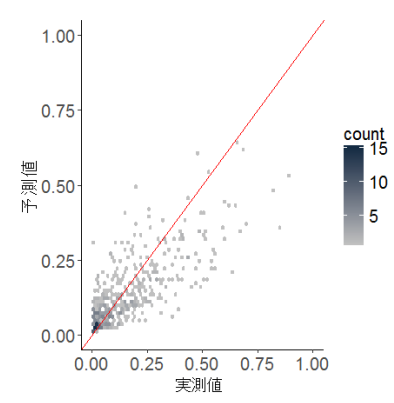


図3 AFLR₂₀₁₅の予測値と実測値の比較

3. 結果と考察

図1から図3に、本研究で構築された最適モデルAとTの予測汎化性能を評価した結果を示す。図1は、最適モデルAにより予測されたAFL₂₀₁₅と、農林業センサによる実測値を示している。実測値が0から220の広い範囲に分散している一方で、予測値は0から120の狭い範囲に分散しており、相関係数は0.81となった。図2は、最適モデルTにより予測されたTFL₂₀₁₅と、農林業センサによる実測値を示している。TFL₂₀₁₅の相関係数は0.96となり、AFL₂₀₁₅よりも広い予測範囲において予測性能が高いことが分かった。図3は、AFLR₂₀₁₅の予測結果と農林業センサによる実測値を示している。AFLR₂₀₁₅の相関係数は0.76となり、前年度までの説明変数から5年後の耕作放棄地面積率を一定程度、予測できたといえる。

表2にモデルA、表3にモデルTで重要だと特定された耕作放棄要因、農地縮小要因の説明変数を示す。共に、農家の規模が重要であると特定された一方で、耕作放棄要因では県別での傾向や所有機械台数を表す説明変数が重要であることが特定された。農地縮小要因では、女性の労働力を表す説明変数が重要であることが特定された。これらの要因分析の結果は先行研究の知見をより精緻に特定できたといえる。

4. 今後の課題

図4に、北陸3県におけるAFLR₂₀₁₅の予測値と実測値との誤差を表したマップを示す。富山市、金沢市、福井市などの都市近郊などの平野部では予測精度が高い一方、アルプス付近の山岳地帯や能登半島の豪雪地帯、森林部などでは過小評価されている地域が多く見られた。この結果から予測モデルの精度を向上させる可能性がある、積雪量や降雨情報、平均気温などの変数を模索し、追加検討する。また本プロセスが北陸地方以外の他地域においても移転できるか検証する必要がある。

参考文献

- [松井 2014] 松井孝典, 宇賀田徹, 町村尚: 機械学習アルゴリズムによる耕作放棄の要因分析および予測モデルの開発, 土木学会論文集 G (環境), vol. 70, No.6 (環境システム研究論文集第42巻), II_131-II_139, 2014.
- [農林水産省 2017] 農林水産省: 農林業センサ, <<http://www.maff.go.jp/j/tokei/census/afc/>>, 2017.02 参照.
- [国土交通省 2017] 国土交通省国土政策局国土情報課: 国土数値情報ダウンロードサービス, <<http://nlftp.mlit.go.jp/ksj/>>, 2017.02 参照.

表2 重要度が上位の耕作放棄要因

順位	説明変数	相対重要度
1	販売農家で所有経営耕地面積の規模が0.3~0.5 [ha] の農家数 [戸]	100
2	都道府県	41
3	勾配: 販売農家が所有するトラクターの台数 [台]	23

表3 重要度が上位の農地縮小要因

順位	説明変数	相対重要度
1	販売農家で所有経営耕地面積の規模が2.0~3.0 [ha] の農家数 [戸]	100
2	販売農家で所有経営耕地面積の規模が10 [ha] 以上の農家数 [戸]	76
3	販売農家で30~34歳の女性農業就業者数 [人]	42

AFLR₂₀₁₅の予測値と実測値との誤差

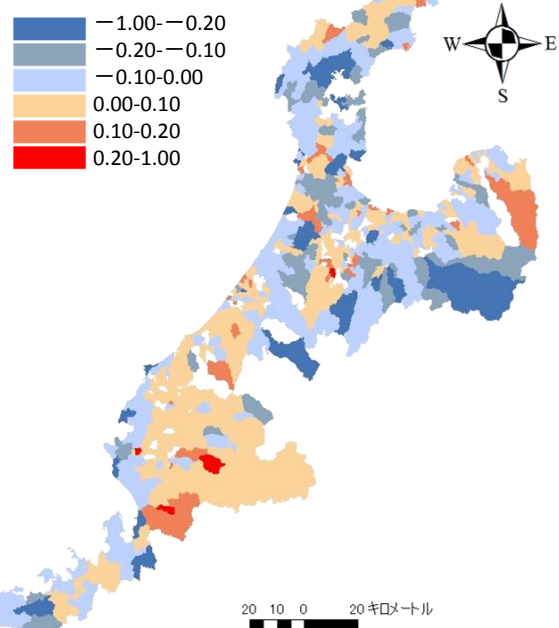


図4 AFLR₂₀₁₅の予測汎化誤差の空間分布