

D3 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

地球循環共生工学領域 08E13036 左右田 篤 (Atsushi SOUDA)

Abstract: 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, I 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.

Keywords: cultivated land abandonment, machine learning, census of agriculture and forestry

1. 序論

日本では耕作放棄が増加しており、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。これに対して先行研究では¹⁾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

(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) 分析用データの収集

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²⁾のうち販売農家の営農情報と、国土数値情報³⁾の5次メッシュスケールでの標高と傾斜を使用した。

(3) 将来予測モデルの設計

式(1)により*i*年の耕作放棄地面積率 $AFLR_i$ を予測した。

$$AFLR_i = \frac{AFL_j + \Delta AFL_{i,j}}{(TFL_j + \Delta TFL_{i,j} + AFL_j + \Delta AFL_{i,j})} \quad (1)$$

ここで AFL_j および TFL_j は*j*年の耕作放棄地面積 [ha]、経営耕地面積 [ha]を表す。 $\Delta AFL_{i,j}$ と $\Delta TFL_{i,j}$ はそれぞれ、*j*年から*i*年までの耕作放棄地面積の変化量 [ha]と経営耕地面積の変化量 [ha]を表す。*j*年は*i*年の5年前を表す。本研究では最終的に $AFLR_i$ を予測するために、 $\Delta AFL_{i,j}$ を予測するモデルA、 $\Delta TFL_{i,j}$ を予測するモデルTを、式(2)と式(3)のようにそれぞれ構築した。

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

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

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された販売農家の*i*年、*j*年、*k*年における共通項目、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、最小値を採用した。これらの説明変数は、「労働力要因」、「経済要因」、「農業経営要因」、「地理要因」の大きく分けて4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、最終的に計159の説明変数が分析用として採用された。

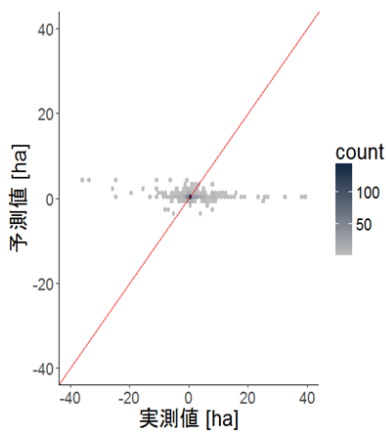


図1 $\Delta AFL_{2015,2010}$ の予測値

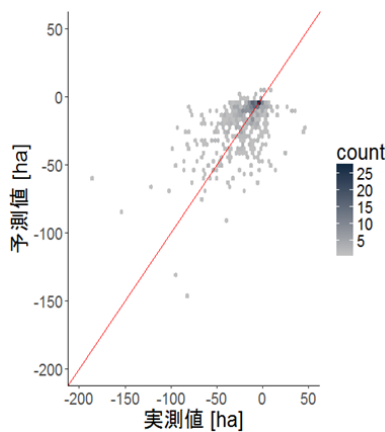


図2 $\Delta TFL_{2015,2010}$ の予測値

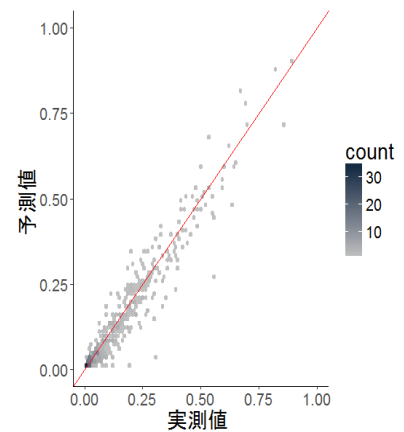


図3 $AFLR_{2015}$ の予測値

(4) 確率勾配ブースティングの適用

式 (2) の f_A , 式 (3) の f_T 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 分割交差確認を行うことで、モデルの学習段階における汎化誤差を評価し、 10×10 の入れ子交差確認によるグリッドサーチにより決定したベストパラメーターでモデルを構築した。

R ver.3.31 の上で `xgboost` と `caret` のパッケージを用いた。

3. 予測汎化誤差の検証結果

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 $\Delta AFL_{2015,2010}$ と、農林業センサによる実測値を示している。実測値が-40から40の広い範囲に分散している一方で、予測値は-4から4の狭い範囲に分散しており、相関係数は-0.21と低い値となった。図2は、最適モデル T により予測された $\Delta TFL_{2015,2010}$ と、農林業センサによる実測値を示している。 $\Delta TFL_{2015,2010}$ の相関係数は0.52となり、 $\Delta AFL_{2015,2010}$ 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、 $AFLR_{2015}$ の予測結果と農林業センサによる実測値を示している。 $AFLR_{2015}$ の相関係数は0.96を記録し、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高精度に予測できた。図4に、北陸3県における $AFLR_{2015}$ の予測マップを示す。

4. 考察と今後の課題

図1と図2に示すとおり、実測値の密度が高いところに予測値が集中してしまった。経営耕地面積と耕作放棄面積の変化には都市化による圧力も影響するため、今後の課題として都市計画区分による開発圧力を説明変数として追加することを検討する。また本モデル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

- 1) 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.
- 2) 農林水産省: 農林業センサ, <<http://www.maff.go.jp/j/tokei/census/afc/>>, 2017年2月参照.
- 3) 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年2月参照.

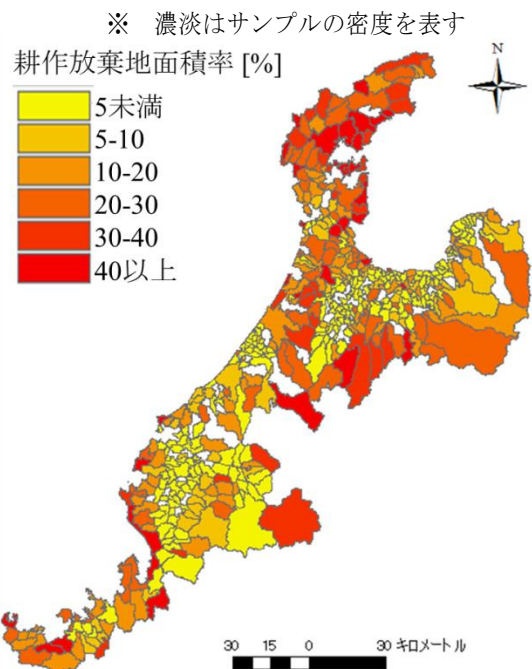


図4 $AFLR_{2015}$ の予測マップ