

要因同定型の機械学習アルゴリズムによる耕作放棄リスク予測

A projection of farmland abandonment and detection of risk factors by machine learning algorithm

松井孝典^{*1} 宇賀田徹^{*1} 町村尚^{*1}
MATSUI Takanori UGATA Tetsu MACHIMURA Takashi

^{*1} 大阪大学大学院工学研究科
Graduate school of engineering, Osaka university

Farmland abandonment has been increasing rapidly in recent years and it causes degradation of biodiversity and decreases ecosystem services from agro ecosystem. Thus, it is required to determine the factors increasing cultivation abandonment and project the future level of abandonment in regional scale. We tried to develop risk factor analysis and prediction model of abandonment of farmland in regional scale. We used census of agriculture and forestry distributed by Ministry of Agriculture, Forestry and Fisheries of Japan. And to identify the structure of the cultivation abandonment from the census data, we employed machine learning algorithms; Random Forests (RF). The result shows that RF could detect major factors under regional specific situations.

1. はじめに

日本では農業従事者の高齢化や労働力不足、農作物の価格低迷による収益性の低下などを背景として、耕作放棄地が増加し続けている。耕作放棄の進行は農業および農村の持つ多面的機能の低下を招き、生態系サービスの供給力を減少させるため、早期の予測と診断を行う必要がある。しかし、一般的に耕作放棄の要因は地域が固有にもつ農業の事情によってそれぞれ異なるため、すべての地域にユニバーサルに適用可能なモデルで要因を説明することは困難を伴う。そこで、本研究では、耕作放棄対策の支援情報を提供すべく、これまで実施してきた、任意の地域においてその地域特有の耕作放棄を促進する重要な要因の特定と、高精度な将来予測を両立する予測モデルを自動生成するプロセスの開発を基に [松井 2014], [宇賀田 2013], モデルの予測精度の向上を行うことを目的とする。

2. 分析方法

一般にモデル生成のプロセスは、(1) モデル生成用データの選定、(2) 応答変数および説明変数の設計、(3) モデル構造の選定、(4) モデルチューニングとバリデーションの順に行われるが、本研究では、これらのプロセスを簡易な演算手順として定式化することで、そのプロセスに従えば様々な地域で固有の状況に応じた耕作放棄要因の特定と耕作放棄量の精度よい将来予測ができるモデルの生成を可能とすることをめざしている。

2.1 地域別の耕作放棄に関連するデータベースの構築

モデルの生成に用いるデータは、[農林水産省 2015] によって公開されている「農林業センサス—農業編—」とした。農業センサスとは、我が国農林業の生産構造、就業構造を明らかにするとともに、農山村の実態を総合的に把握し、農林行政の企画・立案・推進のための基礎資料を作成し、提供することを目的に 5 年ごとに行われる調査である。(1) 経営の態様、(2) 農業労働及び林業労働、(3) 耕地及びその他の土地、(4) 家畜及び蚕、(5) 農業用の機械及び施設、(6)

農業生産物、(7) 農作業、(8) 山林、(9) 育林及び素材生産、(10) その他農林業経営体の現況を把握するために必要な事項の 카테고리 からなる項目を保持した統計表である。1904~2015 年の累年データが完備され、我が国の農業の構造を把握する上で大変貴重な基礎データである。

このうち現時点で利用可能な村落単位の空間解像度で入手可能な耕作放棄に関連するデータとして、2000、2005、2010 年の 3 つの時点の農林業センサスを用いた。なお、先行研究と比較可能となるように、サンプルケースの地域は奈良県 (N=161) とした。

2.2 応答変数の設計と素性の抽出

応答変数は市区町村の面積や農業規模に依存せずに耕作放棄の進行度を表す指標として、耕作放棄地面積を総耕作地面積で除した耕作放棄地割合 R とした。次にこれを説明する素性 x の候補は、生データから得られる約 600 の変数を農業センサスの地域の面積や農業規模によらない変数に標準化した 118 変数を生成した。

[松井 2014] では、この 118 の素性について 2005 年のデータを学習用に利用し、そのモデルで 2010 年のデータの素性の値から耕作放棄量を予測し、実測値と比較して精度を検証しており、2 時点でのデータのみを利用している。これに対して今回は 2000、2005、2010 年の 3 時点のデータを利用して

$$\text{学習: } R_{2005} = f(x_{2005}, \nabla x_{2000-2005}) \quad (1)$$

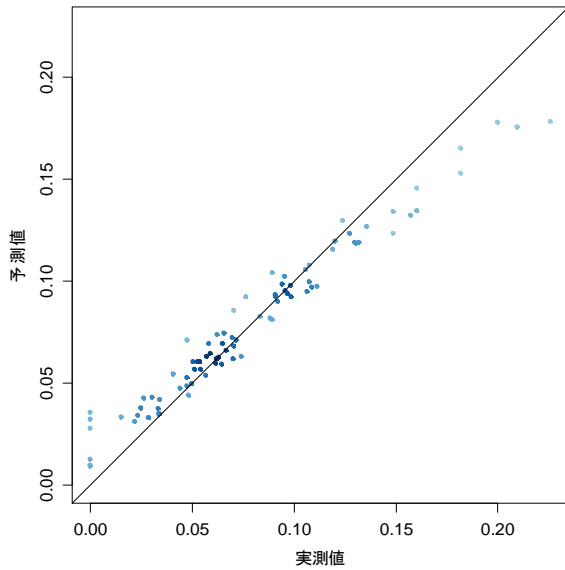
$$\text{予測: } R_{2010} = f(x_{2010}, \nabla x_{2005-2010}) \quad (2)$$

のように勾配情報を追加し、236 の素性ベクトルに拡張した。

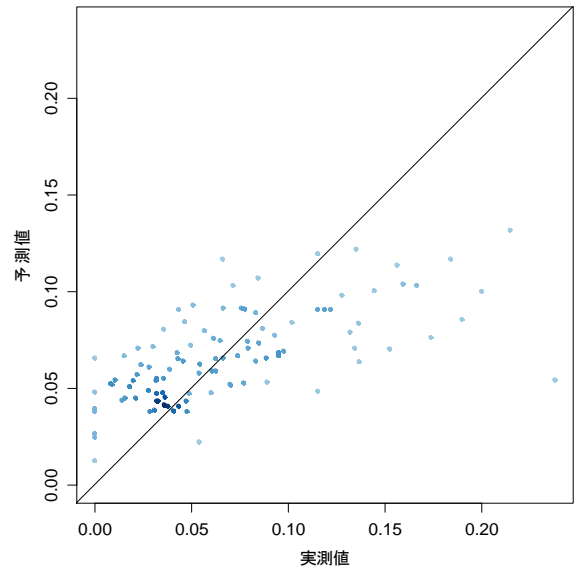
2.3 予測モデルの構築

耕作放棄が進展すると予測された場合には、その要因を緩和するための政策的な支援を行う必要があるため、要因が明確に特定される必要がある。今回の f には、先行研究で予測の精度とテラーメードな特定の要因の両立する可能性を示した Random Forest を用いた。チューニングでは、①モデルに採用する素性ベクトルの次元数、②ランダムサンプリングする素性数、③末端ノードの最少個体数の 3 つのパラメータを二乗誤差基準により 10-分割交差検定で最適化した。

連絡先: 松井孝典, 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, matsui@see.eng.osaka-u.ac.jp



a) 2005年のデータに対する学習



b) 2010年のデータに対する予測

図1 予測モデルの学習および予測結果

3. 結果および考察

3.1 識別器の性能評価

まず、図1にモデルの学習および予測の結果を示す。図1 a) が2005年のデータを用いたモデル学習段階、図1 b) が学習後のモデルで2010年のデータが再現できるかを示したものである。濃淡は個体の密度を反映している。2000-2005年のデータによる学習段階では、交差検定とグリッドサーチによるパラメータ最適化後の学習モデルの精度は R-squared: 0.981 となり、0~0.2 付近で散布する耕作放棄地率に対してどのレベルでも概ね良好なフィットを示した。一方で、2010年のデータの再現性については、R-squared: 0.781 の精度があるものの、局所的な傾向としては、耕作放棄地率が 0~0.05 付近の低耕作放棄レベルで過大評価、0.1~0.2 付近の高耕作放棄レベルで過小評価の傾向を示した。

3.2 特定された耕作放棄要因

表1に、Random Forest から得られる相対重要度の順に、特定された耕作放棄要因および耕作放棄量との相関係数の符号を示した。負の符号を持つ要因は耕作放棄が増加することを緩和する要因であり、「R3:▽専従者なし準専従者は女子だけ農家率」、「R8:農産物販売額 500 万円以上農家率」、「R13:▽いも類作付農家率」、「R14:▽果樹園のある農家率」の4つの要因が抽出された。この結果は、労働力がある程度確保される、ある一定程度の経営母体を有している、という一般的な営農に関係する要因であるが、これに加えて大和いも、イチゴ、柿などの奈良県を代表する特産品を栽培するかどうか耕作放棄リスクの緩和に寄与する可能性があるとして解釈できる。

4. 今後の課題

2017年3月頃に農林業センサスが公開される予定であるため、式(1)、(2)を以下に更新すると共に、新たな要因同定型の機械学習アルゴリズムの開発を試みる。

表1 モデル構築に使用された素性と相対重要度

相対重要度	素性	相関
100	畑のある農家率	+
71	畑のある面積率	+
67	▽ 専従者なし準専従者は女子だけ農家率	-
64	専従者なし準専従者もない農家率	+
56	豆類作付農家率	+
56	専従者なし準専従者は女子だけ農家率	+
53	田面積率	+
52	農産物販売額500万円以上農家率	-
48	コンバイン利用農家率	+
43	▽ 樹園地面積率	+
42	▽ 田面積率	+
40	▽ コンバイン利用農家率	+
39	いも類作付農家率	-
38	▽ 樹園地のある農家率	-
36	▽ 乗用型トラクター利用農家率	+
34	▽ 基幹的農業従事者男性平均年齢	+

備考：▽は予測年までの5年前からの変化の勾配、無印は予測年の状態量を表す

$$\text{学習: } R_{2010} = f(x_{2005}, \nabla x_{2000-2005}) \quad (3)$$

$$\text{予測: } R_{2015} = f(x_{2010}, \nabla x_{2005-2010}) \quad (4)$$

参考文献

- [宇賀田 2014] 宇賀田徹, 松井孝典, 町村尚:奈良県における耕作放棄の要因分析および予測モデルの開発, 土木学会第41回環境システム研究論文発表会講演集, pp. 39-44, 2013.
- [農林水産省 2015] 農林水産省:農林業センサス, <<http://www.maff.go.jp/j/tokei/census/afc/>>, 2016.03 参照.
- [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, pp. II_131-140, 2014.