

科学研究費助成事業 研究成果報告書

令和 4 年 6 月 8 日現在

機関番号：12102

研究種目：挑戦的研究（萌芽）

研究期間：2019～2021

課題番号：19K21686

研究課題名（和文）経済学の実証研究における衛星画像と機械学習の応用 アフリカの開発政策を事例に

研究課題名（英文）Satellite Imagery and Machine Learning in Economics: A Case Study of Development Policies in Africa

研究代表者

牛島 光一（USHIJIMA, KOICHI）

筑波大学・システム情報系・助教

研究者番号：80707901

交付決定額（研究期間全体）：（直接経費） 5,000,000円

研究成果の概要（和文）：本研究の目的は、衛星画像を用いて行う機械学習が、サブサハラ・アフリカのような統計情報の不足する地域の情報収集のための強力なサポートツールになることを示すことである。入力（衛星画像）に対応する出力（経済的情報）がある時期の情報を教師データとし、出力データのない時期のための予測モデルを構築した。さらに他年度のデータから、予測値パネルデータを構築し、民族優遇に関する仮説を検証した。道路の量、質ともに大統領と同じ民族への優遇があることが確認できた。また、全道路のデータを用いた舗装の有無に関する機械学習の結果、偽陰性率はやや高いものの、偽陽性率のかなり低い予測モデルを構築した。

研究成果の学術的意義や社会的意義

本研究の成果は、途上国のデータが欠損する地域への応用可能性を示したのみならず、我が国にも重要な社会的意義がある。人口減少により政策予算が限られていく我が国にとって、地域間・地域内の輸送コストを管理するための道路資本ストックの維持は最重要課題の一つである。本研究は無料公開されている衛星画像を用いることで、道路の質に関する大規模調査情報からパネルデータを構築できる可能性を示すことができた。大規模道路インフラ調査のためには莫大な費用が必要になるが、衛星画像から一定水準以上の精度で予測値をもとめることで、効率的に道路資本ストックの管理を行えるようになる。

研究成果の概要（英文）：This study shows that machine learning performed with satellite imagery can be a powerful support tool for collecting information in regions such as sub-Saharan Africa, where statistical data is scarce. We used information from areas with output (economic information) corresponding to the input (satellite imagery) as teacher data to build a forecasting model for periods with no output data. We also used data from other years to construct predictive value panel data to test hypotheses about ethnic favoritism. Both the quantity and quality of roads were found to favor the same ethnic groups as the President. Machine learning results on pavement availability using data for all roads resulted in a predictive model with a rather low false positive rate, although the false negative rate was somewhat high.

研究分野：開発経済学

キーワード：機械学習 衛星画像 統計的因果推論 サブサハラ・アフリカ

1. 研究開始当初の背景

2015年に「持続可能な開発目標 SDGs」が採択され、途上国への有効な開発支援政策のあり方を再検証する機運が高まっている。多くの途上国や開発支援組織は厳しい予算制約に直面しており、より有効性の高い開発支援政策を選択していくために、より質の高いエビデンスの蓄積が不可欠である。しかしながら、所得水準の低い途上国ほど統計情報の質は低く、信頼性の高いエビデンスの蓄積が困難という問題が認識されていた。近年、この問題の解決策として、衛星画像から経済的情報を生成する研究手法が注目されるようになってきている。例えば、Henderson et al. (2012 Am Econ Rev) では、national accounts の精度が低い地域に対して、衛星から観察した夜の光の強さを使い経済成長の良い指標としている。衛星画像を用いた研究が増えつつあるが、途上国が直面する様々な課題解決のための情報を衛星画像から作成している研究は依然として少ない。

2. 研究の目的

本研究の目的は、衛星画像を用いた機械学習の手法が、サブサハラ・アフリカのような統計情報の不足する地域の情報収集のための強力なサポートツールになることを示すことである。具体的には、入力（衛星画像）に対応する出力（経済的情報）がある時期の情報を教師データとし、出力データのない時期のための予測モデルを構築する。また、この手続きを「時期」から「地域」に応用することで、転移学習という手法に応用できる可能性がある。衛星画像は地域によって入力データの分布・性質が異なることがあるが、転移学習は入力データの分布・性質の違いを補正して高精度な予測を実現する手法である。すなわち、例えば、サブサハラ・アフリカのある地域において社会経済的な情報に関する十分な教師データがあるならば、近隣の諸国、もしくは、より広い範囲の国々の社会経済的な状況を高精度に予測することができる可能性がある。

3. 研究の方法

本研究は衛星画像を用いた機械学習が途上国における社会経済的な情報を補完することを示すために、ケニア全土における道路の質情報（予測対象）と衛星画像（入力）を教師データとし、異なる年の道路の状況を予測する。予測されたデータを用いた応用例として、大統領の民族優遇の問題を考える。

手順

(1) 衛星画像から空間情報を作成する。衛星画像はラスタデータというピクセルの集合体であるため、道路の空間情報に対応した画像を抽出する。(2) ケニアの Kenya National Land Commission から提供いただいた全土の道路の質情報と道路ベクターデータを教師データとし道路の質の予測モデルを構築する。(3) 予測モデルを用いて道路の質情報を外挿する。このとき予測モデル構築に用いた衛星画像と外挿対象の地域の衛星画像の差異を、転移学習の技術を用いて補正して高精度な外挿を実現する。さらに、外挿する地域の追加的な情報をもとに予測の精度を向上させる。(4) 構築されたデータを用いた実証研究を行う。

4. 研究成果

ケニアを対象とした 2 つの研究を進めた。(A) 道路の質と民族的依怙贖員 (ethnic favoritism) の因果関係について、(B) 大規模道路データの機械学習についてである。

(A) 道路の質と民族的依怙贖員 (ethnic favoritism)

この研究では、大統領の民族的依怙贖員のために、民族の居住地によって道路状況(量、

質)に差があるのかどうかについて調べた。因果関係を調べるために2つの識別戦略を採用した。第1は、空間的回帰不連続デザイン(SRD)、第2は空間的回帰不連続デザインを用いた差分の差分法(SRD-DID)である。

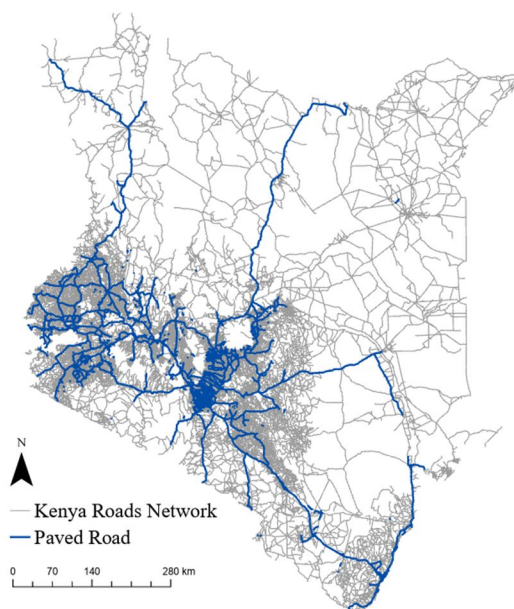
ケニアは大英帝国植民地時代の統治政策の影響により、現在の行政界の境界が民族ごとの居住地境界と一致している。つまり、民族的依怙鬮肩があるのであれば、行政界の境界線付近でさえ、道路の整備状況に差が観察される可能性がある。優遇される民族は2つあり、1978年から2002年まで大統領だったダニエル・アラップ・モイのカレンジン族と2002年以降大統領を排出しているキクユ族である。

SRDモデルでの分析の結果、道路の舗装状況、舗装道路の質、未舗装道路の質、は全て2つの民族のほうが統計的に有意に優遇されている結果となった。道路ネットワークよりも民族優遇のほうが優先されていたと考えられる状況である。SRD-DIDモデルの分析では、(B)の方法(機械学習)で境界付近の舗装道路に関して予測値パネルデータを構築した。1992年の大統領選挙の情報を利用して、選挙の際に大統領を支持した地域は他の選挙区とくらべて道路の量が多い、ただし、大統領と同じ民族でないと道路の質は悪い、という2つの仮説について、分析を行った。キクユ族の居住地域の道路の質には改善傾向はなく、カレンジン族の居住地域の道路の質は年々改善する傾向がみられた。1992年の選挙でモイ大統領が大勝したカレンジン族以外が住む地域(カレンジン族、キクユ族以外のケニア有力民族が住む地域)では、道路の質に改善傾向はなく、接戦で勝利した地域(カレンジン族、キクユ族以外のケニア非有力民族が住む地域)では道路の質に改善傾向が見られた。選挙の支持のために予算配分が行われたことが伺われる。

(B)大規模道路データの機械学習

2つの機械学習とパネルデータの構築を行った。(A)の研究で用いる舗装道路の質パネルデータを構築した。まず、行政界境界付近の舗装道路に対応する衛星画像を収集した。その画像と道路空間情報を重ね合わせることで、舗装道路の位置に対応する画像を抽出し、機械学習によって予測モデルを構築した。教師データと異なる年の画像を収集し、予測モデルを用いて予測値パネルデータを構築した。

舗装道路の質が低い場合は、未舗装道路と同じように通行が不便になるため、とは異なり、舗装道路かどうかに関する予測モデルを構築した。道路は舗装・未舗装あわせて16万本あり(図を参照)それに対応した年度の画像を収集し機械学習を行った。畳み込みニューラルネットワークによって予測モデルを構築した。予測モデルの精度は、正解率(Accuracy)89.2%、偽陰性が44%、偽陽性が7%と、未舗装道路を舗装道路と間違える確率は極めて低く、舗装道路を未舗装道路と間違える可能性のやや高い予測モデルとなっている。これは、舗装道路の舗装が剥がれているケース(=実質的には未舗装道路)が少ない状況を鑑みると決して悪い数字ではない。現在は、予測モデルの向上に取り組みつつ、教師データとは異なる年度の衛星画像を収集し、この予測モデルを用いて道路の舗装状況に関する予測値パネルデータの構築を進めている。



図：ケニアの道路ネットワーク(2008)

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計0件

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	日野 英逸 (Hino Hideitsu) (10580079)	統計数理研究所・モデリング研究系・教授 (62603)	
研究分担者	小西 祥文 (Konishi Yoshifumi) (40597655)	慶應義塾大学・経済学部・教授 (32612)	
研究分担者	木島 陽子 (Kijima Yoko) (70401718)	政策研究大学院大学・政策研究科・教授 (12703)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関