

開発経済学におけるビッグデータの活用

岡部 美砂*

Applications of Big Data to Development Economics

Misa OKABE

要旨：情報通信技術の進展により多様なデジタル化データが日々膨大に生成されている。それらいわゆるビッグデータを用いた研究が、開発経済学のフロンティアを拡大しつつある。本稿は、携帯電話の通信記録および遠隔計測データを用いた研究をサーベイし、開発経済学でビッグデータがどのように利用されているか、どのような成果を得ているのかを紹介し、その課題を考察する。

Abstract :

A vast variety of digitization data, what is termed Big Data, is generated enormously every day with the development of the information and communication technology. Studies by using Big Data are expanding the frontier of development economics. We review studies by using cell-phone call details records and remote sensing data to introduce how Big Data is used, what kind of result we got in development economics. Also, we consider the challenges of utilizing Big Data for development economics.

キーワード：開発経済学、ビッグデータ、CDR、リモートセンシングデータ、サテライトデータ

はじめに

情報通信技術の進展によって、日々膨大なデータが生成されている。それらいわゆるビッグデータを用いた研究が経済学のフロンティアを拡大しつつある。特に近年、経済学の実証分析では、マイクロデータの利用拡大、フィールド実験や経済実験など新しいデータの開発と、その分析手法の開発が急速に進展しており、経済学における実証分析の比重が高まりつつある¹⁾。ビッグデータは、実証研究を重視してきた経済学の進展をさらに押し進める要素の一つでもある。

また、ビッグデータは、従来の統計ではカバー

されなかった分野のデータを利用可能にすることから、応用経済学の多くの分野で利用され始めている。特に、データの利用可能性の制限が多かった開発経済学では、ビッグデータの特徴を利用し、多様な研究が進められている。先進国に比べて、途上国では物的・人的資源の不足から経済分析に必要な基本的なデータの整備が遅れていることが多く、統計的な手法を用いる際の制約となっている。特に、近年の経済学の実証研究で必要となっているマイクロデータは、最近になってようやく収集・公開され始めたケースが多く、分析の制約となっている。いわゆるビッグデータは、人々や企業の行動データを網羅的に収集・記録したも

*和歌山大学経済学部准教授

1) Hamermesh (2013)によると、ここ60年の経済学のトップジャーナルに掲載された論文の傾向から、理論研究の論文が減少傾向にある一方、一次データやフィールド実験による実証研究の論文数が急増している。

なので、マイクロレベルデータの不足を補うものとしても活用が進んでいる。

本稿では、開発経済学においてビッグデータがどのように活用され研究が進められているのかを紹介したい。特に途上国でも広く使われている携帯電話の通信記録を用いた研究と、人工衛星から地上を計測した画像を用いた研究に焦点をあてて、これらのビッグデータによって、開発経済学ではどのような研究が新しく可能になっているのか、またそれによってどのような発展と課題がもたらされるのかを考察する。

1. ビッグデータ分析が 経済学に与える影響

1.1 ビッグデータとは

いわゆるビッグデータといわれるものには、携帯電話やインターネットの通信記録、コンビニエンスストアなどの小売店での販売記録、電子マネー取引、電子化されたビジネス業務の記録、ICカードによる公共交通の利用記録、さらには人工衛星など遠隔から計測した画像などの記録等、多種多様な電子化されたデータが含まれている。ビッグデータを定義する言葉として、Laney (2001) が使った「3つのV (Volume, Variety and Velocity)」がよく使われており、ビッグデータとは、膨大な情報量で、多様な種類、そしてリアルタイムで収集されるデータであることを特徴とするデジタル化されたデータの総称であると言える。また、これらの膨大なデータを収集・記録し保存するためのストレージシステムの急速な進化がビッグデータを支えている。

近年、これらのビッグデータの社会・経済への応用が目ざされている。例えば、ICカードを用いた乗客の行動予測からバスの運行の最適化を測ったり、データ化された医療記録から効果的な医療健康政策を立案したりするなど、様々なデータが、人々の行動パターンを予測し、社会や経済のシステムの効率化を図ることに利用されて始めている。

1.2 経済学における活用

経済学で実証分析に使用するデータは、主に公的機関が発行する統計データ、および公的機関または民間業者が実施したアンケート調査が中心であり、近年では、フィールド実験によって研究者が収集したデータも使用されている。それら従来経済学で用いられてきたデータと、ビッグデータとの大きな違いは、従来のデータは基本的に経済学分析を目的として調査収集されたものであるのに対し、ビッグデータは収集段階では分析が目的ではないという点にある。つまり、ビッグデータは、やみくもにすべての対象の情報が記録されたデータであるとも言える。ゆえに、ビッグデータを経済学に応用するためには、いかに経済学で仮定するフレームワークにあったデータを見つけるのか、またはどのように加工すればいいのか、その方法を吟味する段階が必要になる。

その一方、ビッグデータは、Hilbert and Lopez (2016) が述べているように、データ量が多いことは変数間の因果関係や、説明変数の有意性を正確に見出すことが可能になるという利点も持っている。また、これまで収集が困難であったデータを利用可能にするということはビッグデータの大きな利点である。例えば、マクロ経済学では価格調整の時間は重要な関心事であるが、全ての財・サービスの価格データを収集して市場価格の変化をとらえることは困難であった。現在では、その価格データを小売店の販売データ (POS) や、インターネットで収集するオンライン価格を用いて分析する試みが行われている²⁾。経済主体の行動のメカニズムを解明し、その相互作用から経済全体を読み解く経済学にとって、個人や企業の行動を追える大量のデータを使用できるということは大きなメリットであると言える。

2. 開発経済学で 使われ始めるビッグデータ

開発経済学は、ビッグデータの特性を大いに活用できる分野である。途上国ではこれまでデータ

2) 例えば、一橋大学物価研究センター (<http://www.ier.hit-u.ac.jp/~ifd/purposeplan.html>) や、マサチューセッツ工科大学の Billion Prices Project (<http://bpp.mit.edu/>) などがある。

の利用可能性の制限が大きかった。特に、多様な経済的課題の解決が緊急に必要である最貧国において、基本的なマクロデータであっても長期時系列の使用が困難であったり、統計の正確性・信頼性に問題があったりするケースも見られた。このようなデータ不足の問題を補う試みとして期待できるのが、様々なビッグデータである。近年、国際連合や世界銀行などの国際機関では、ビッグデータを開発途上国の分析や政策立案に活用するためのプロジェクトが進められている³⁾。そこで主に用いられているのは、携帯電話による通信記録である。通話だけではなく、SNS 投稿などの豊富な情報を活用することによって、幅広い研究が行われ、また実際の政策立案に応用されている。また、途上国を対象とした研究では、人工衛星の遠隔計測データも数多く活用されている。特に、夜間光の画像は一般に公開されており、これを経済活動水準の代理変数として用いる研究はすでに数多く蓄積されている。以下では、携帯電話による通信記録と、人工衛星画像データを使用した研究を紹介する。

2.1 携帯電話通信記録

ITU (2016) の推計によると、2016 年時点で、全世界人口の 95% (約 70 億人) が居住する地域で携帯電話が使用可能となっている。携帯電話の加入者数は、発展途上国地域だけでも約 58 億人に達し、図 1 が示すように、欧州地域の加入者数に変化が少ない一方で、特に、アジア太平洋地域、アフリカ地域のように途上国が多い地域でこの 10 年に急速に加入者数が増加している。GSMA (2016) によると、携帯電話の全人口あたり普及率を、実際に携帯電話を使用している人の数 (ユニークユーザー数) で計測すると、2015 年時点で途上国地域全体では 59%、そのうち最も低いアフリカで 43%、アジア太平洋地域では 62%、ラテンアメリカ地域では 68% である。途上国では 10 代までの若年層の占める割合が高いため、アフリカ地域でも成人の半数以上が携帯電話を使用していると考えられ、携帯電話は最も広く普及している通信手段であると言える。

携帯電話による通信記録 (Call Detail/Data Record: CDR) は、通常、携帯電話事業者が管理・保有するデータで、携帯電話サービスの提供や課金のために必要な情報である。CDR には、発信

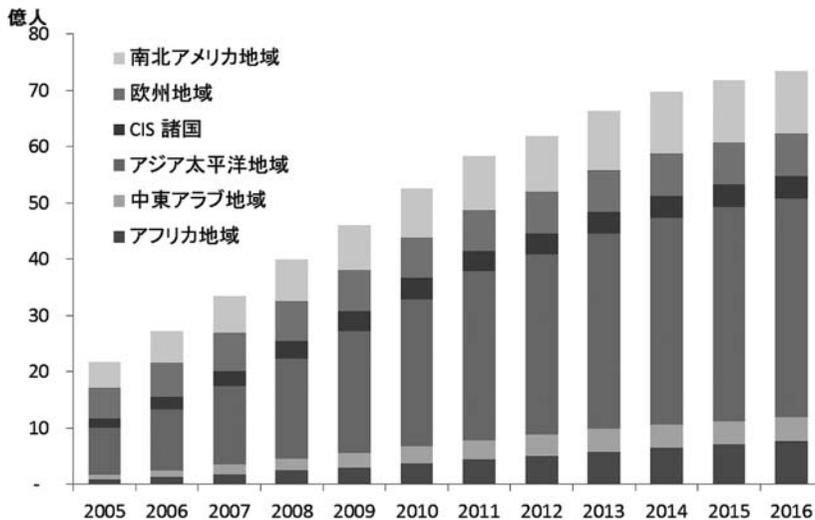


図 1 携帯電話契約者数の推移 (億人)

データ：ITU (2016), end-2016 estimates for key ICT indicators より作成。

注：携帯電話契約者数は一人で複数の契約を行っているケースも含まれており、ユニークユーザー数 (実際の携帯電話使用者数) とは異なる。また、2016 年の数値は推計値である。

3) 例えば、国際連合の Global Pulse Project (<http://www.unglobalpulse.org/>) を参照。

元と発信先の端末の位置、通信日時、時間などが含まれる。また音声通話だけではなく、SNS や MMS を利用したテキストメッセージの送受信やインターネットの通信量も記録されている。これらの膨大かつ広範囲の情報は、途上国を対象とした広範囲の研究にも応用され、従来の統計データやアンケート調査では困難であった、詳細かつ高頻度の分析が行われつつある。以下では、携帯電話の CDR を用いた貧困状況の把握、自然災害・感染症と人々の行動に関する研究の一部を紹介する。

2.1.1 貧困マップの作成

開発経済学の最も重要な課題の一つに、途上国の貧困削減に有効な政策を立案するための貧困のメカニズム解明が挙げられる。そのためにまず必要であるのが、できるだけ詳細かつ最新の貧困の状況を判断できるデータである。従来、政府が実施する国勢調査による生活水準の集計結果によって各国の貧困水準の把握が行われてきた。全数調査によるため正確な貧困状況の把握ができる反面、実施と集計にかかる経済的・時間的費用が大きいため、数年毎のデータしか利用することができないという短所があった。近年になって、CDR を用いて貧困状況（貧困マップ）を把握する試みが行われつつある。Eagle, Macy and Claxton (2010) のイギリス国内の CDR を用いた個人レベルの社会的ネットワークと国勢調査を組み合わせた研究を嚆矢として、途上国においても貧困水準を計測する試みが数多く行われている⁴⁾。

Smith, Mashhadi and Capra (2013) は、二度にわたる内戦のため 90 年代末以降国勢調査が実施されていないコートジボワールを対象に、500 万人の携帯電話加入者の CDR を用いて地域レベルの貧困水準を計測した。彼らは、全国のアンテナ間通信量と通話時間の 1 時間毎の集計データを用

いた通信頻度と時間あたり密度と、貧困水準を計測した既存の指標との相関関係を調べた⁵⁾。分析の結果、同地方内および、また他地方への通信が多いほど社会・経済的な活動が活発で経済的に豊かであり、さらに同地方内の通信の多さの割に他地方への通信が少ない地域ほど貧困度が高いことを明らかにし、携帯電話の通信量を経済活動水準の代理変数として用いることで、地域ごとの貧困度を推計できることを示した。同様に、Pokhriyal, Dong and Govindaraju (2015) は、セネガルの省別の貧困水準を CDR によって計測できるか検証した。彼らは CDR をネットワークと見なし、ネットワークの中心になる地域の貧困度が低いことを見出し、そのデータと貧困指標を用いて群・区レベルの詳細な貧困マップを作成した。CDR によって、細分化された地区ごとの貧困度の推定を可能とすることで、より有効な貧困政策の立案に貢献できることを示した。

さらに、Blumenstock, Cadamuro and On (2015) はルワンダを対象に、CDR と、加入者の資産や住居などの経済状況に関するアンケート調査を組み合わせることで、携帯電話の使用履歴から個人の経済状況を予測できることを示した。この結果から携帯電話加入者履歴を用いた小区域 (microregion) を単位とする詳細な貧困マップを作成している。また、Soto, Frias-Martinez, Virseda and Frias-Martinez (2011) は、ラテンアメリカの主要都市を対象として、CDR による経済状況の判別は、家計調査を基にすると 80% 以上の正確な識別が可能となることを検証した⁶⁾。このように、これまで詳細かつ最新の貧困状況の把握が困難であった途上国においても、CDR を用いることで、より詳細かつ最新の貧困状況を正確に把握することが可能となってきている。

4) Eagle, Macy and Claxton (2010) は、CDR を用いて、社会的なネットワークの多様性と、その個人が住む地域の経済厚生との正の相関を見出した。

5) 既存の貧困水準をあらわす指標は、オックスフォード大学の貧困・人間開発イニシアティブ (OPHI) の「多次元貧困インデックス (multidimensional Poverty Index (<http://www.ophi.org.uk/policy/multidimensional-poverty-index/>)) が用いられた。

6) 同じく CDR と加入者を対象としたアンケート調査を組み合わせ、Sundsøy, (2016) は同地域内での通信量、SMS の受信数に対する発信数、そして通信相手先数と、非識字の相関を見出し、非識字率マップを作成する方法を提案している。

2.1.2 自然災害および感染症拡大と個人の行動パターン

洪水や台風、地震・津波などの自然災害による損害は途上国ほどより深刻となる。また、感染症の発生率も拡大による被害も、やはり途上国でより深刻な問題となっている。自然災害も感染症も貧困リスクを高める大きな要因となる。ゆえに、自然災害による被害の軽減や、感染症拡大の防止のために、人々の行動メカニズムを理解し、限られた人的物的資源を有効に配分して、貧困リスクを軽減する方法を考えるのは開発経済学の課題でもある。しかし、災害や感染症などの発生直後の人々の行動を即時かつ正確に把握することは特に途上国では困難を伴う。そこで、携帯電話の CDR をいることによって、発生当時の人々の行動パターンを観察する方法を開発する試みが行われている。

ハイチでは 2010 年に発生した大地震、および同年のコレラの感染拡大による甚大な被害が発生した。Bengtsson, Lu, Thorson, Garfield and Schreeb (2011) は携帯電話の SIM カードの位置情報を用いて、災害の前後の人々の移動パターンを分析し、首都ポルトープランスから市外へ避難する人の移動は地震後の早い段階で生じ、地震後 19 日以内に居住人口の 20% が減少したと推定している。この分析をさらに拡張した Lu, Bengtsson and Holme (2012) は、地震後 3 日間は首都に留まる割合が大きいが、それ以後の移動が急速に生じること、および早い段階で移動した人々は、親族など社会的つながりの強い人の居住する場所へ移動するというパターンがあることを見出した。

Ghurye, Krings and Frias-Martinez (2016) は、2012 年にルワンダで起きた大規模な洪水によって変わる人々の行動パターンを、洪水前の CDR を用いた機械学習によって導出した平常時の行動パターンと比較して分析を行っている。その結果、人々の移動は洪水発生から 3 週間までが最も多く、次第に減少するものの、10 週間が過ぎても平常時を上回るペースで流出することが分かっ

た。これは、人々が元の居住地に戻るまで 2 カ月以上かかるかもしれない可能性があるとしている。さらに、主な移動先は地方で、少数の知り合いが居住する場所であることを見出している。

一方、感染症の拡大の分析について、Wesolowski, et al. (2012) は、ケニアの約 1500 万人の CDR から人々の移動パターンを読み取り、マラリア伝染モデルと組み合わせて、マラリアの拡大ルートマップを作成し、地域別にマラリア拡大のリスクを判別することを可能にした⁷⁾。同じく、CDR を用いて Frias-Martinez, Williamson and Frias-Martinez (2011) は、人々の移動パターンに基づいてウイルス感染拡大のモデルを構築し、2009 年にメキシコで発生した新型インフルエンザ (H1N1) の流行に適用して、政府による移動規制のウイルス伝染拡大阻止への効果を測定した。その結果、政府の移動規制措置は、ウイルス拡大のピークを 10% および拡大を 2 日間遅らせる効果があったことを見出した。感染症拡大の予測に CDR を応用することによって、Frias-Martinez, et al. (2011) が指摘するように、従来の感染症伝染モデルで捉えられなかった感染拡大プロセスに内在する複雑性を、個人の行動パターンを組み込むことで捉えられるようになり、より有効な予測モデルを構築することが可能になってきている。途上国で一般的な通信手段となりつつある携帯電話の CDR データは、人々の行動のメカニズムを解明するという開発経済学の基本的な分析手法の発展に貢献している。

2.2 遠隔計測データ

近年になって経済学で使用され始めた遠隔計測 (Remote Sensing) データの主なもの、人工衛星によって測定される地球表面の画像から得られる情報である。表 1 は経済学でよく用いられる人工衛星データを挙げている。これらの人工衛星によって測定された画像には、夜間の光の照度、気候や地形だけでなく、農地の使用状況、建築物の

7) Wesolowski et al. (2012) は、分析では国境を越えた移民や、非季節性のマラリア流行の影響が考慮できていないことに注意すべきと述べている。

表1 経済学でよく用いられる人工衛星データ

観測衛星・データソース	解像度	経済学に応用されるデータの例
ランドサット (Landsat)	30 m	都市開発、森林占有率、天然資源
中分解能撮像分光放射計 (MODIS)	250 m	大気汚染、漁獲量の推計
米国国防省気象衛星計画 OLS 光学センサー、可視赤外面像放射計センサー (DMSP-OLS, VIIR)	>1 km	所得水準、電気利用状況
スペースシャトル地形データ (SRTM)	30 m	標高、地形
Digital Globe 社衛星画像	>1 m	都市開発、森林占有率

出所：Donaldson and Storeygard (2016) より引用し補足して作成した。

数や種類、さらには大気汚染の程度など幅広い種類のデータが含まれる。Donaldson and Storeygard (2016) は、これまで自然科学研究で用いられていた遠隔計測データが社会科学でも広く用いられるようになった背景として、大容量データが一般にも利用可能になったこと、必要な情報を取り出すための多様なアルゴリズムが開発されたこと、クラウドベースでグローバル規模の情報処理が可能になったことを指摘している。

開発経済学でも、遠隔測定データを単体または複数、もしくは他の既存データを組み合わせ、途上国が直面する経済問題や経済現象におけるメカニズム解明のための実証研究に活発に用いられ始めている。特に、遠隔計測データは全世界を網羅し、かつその時系列変化を観測できるため、これまで困難であった途上国地域の様々な変数を大規模なパネルデータとして利用することができるという大きな利点がある。以下では、多様な画像データを用いた研究を、貧困分布の推定、インフラ投資の経済効果、災害・環境汚染・気候変動の経済活動への影響の研究に分け、各分野の研究の一部を紹介する。

2.2.1 夜間の光量による経済活動水準の計測

DMSP-OLS により計測され一般に公開されている夜間の地上の光量データは、地表を 1 km 以下の解像度で記録され、デジタルアーカイブとしては 1992 年以降保管・公開されている (Hender-

son, Storeygard and Weil (2012))。このデータを経済学の実証分析に応用した研究は数多く行われており、それらの研究の主なものは、夜間の光量を経済活動が活発に行われていることの代理変数として、経済成長や経済発展水準を測る指標として用いるものである⁸⁾。

夜間の光量データは、これまで分析が困難であった途上国の都市レベル、地方自治体レベルでの経済活動水準を分析することを可能にした。さらに、人工衛星の計測は継続して行われているので、最新のデータや、時系列データによる分析が可能である。また、Henderson et al. (2012) が指摘するように、特に途上国では都市部と地方の経済活動の差が大きく、政府統計で公表される国レベルの GDP では経済活動水準の実態を把握するのは困難となる。それに対して、夜間の光データの利用は、市区町村レベルのような小さい区分の地域の経済活動水準を推計できることに大きな優位性があるといえる。そのような研究例として、Harari (2016) は、夜間の光量データを用いてインドの都市レベルの経済活動規模を推計し、個人や企業の居住・立地選択に与える影響を空間的一般均衡モデルの枠組みで分析している。さらに、Michalopoulos and Papaioannou (2014) は、夜間の光量データでアフリカの民族単位の経済活動水準を推計して、所属する国(政策・制度)の違いによって、同一民族内でも経済発展水準が異なる

8) 夜間の光量が経済活動水準や経済成長の代理変数として適切であるかどうかの精査も近年多くの研究で行われており、例えば、Henderson et al. (2012)、Chen and Nordhaus (2011) は夜間の光量と GDP 成長に正の相関があることを見出している。

かどうか検証した。その結果、首都圏については民族よりも国の制度・政策の影響が強いが、首都から離れるほどその影響は低下することを見出した。人工衛星画像による夜間の光量を経済水準の代理変数として用いることで、経済成長・経済発展の分析の範囲が大幅に広がっている。

2.2.2 遠隔計測画像を用いた研究

1) 貧困の分布と水準

前節の人工衛星画像による夜間の光量を用いた経済活動水準の測定の限界として、最も貧しい地域間、すなわち光量がゼロである地域間の経済活動水準の差が捉えられないことが挙げられる。それに対して、Blumenstock (2016), Jean, Burke, Xie, Davis, Lobell and Ermon (2016) は、同じ地域の昼間画像から捉えた道路の舗装や屋根の状態と、夜間の光量とを組み合わせて、機械学習によって、最貧地域の経済活動水準を推定している。

人工衛星の高解像度の画像を用いると、民家の数や、屋根の材質や経年変化までとらえることができるので、生活環境を直接観察して所得水準を推定することも可能である。例えば、Marx, Stoker and Suri (2015) は、ケニアの首都ナイロビのスラム地区の住宅投資状況を高解像度画像で分析し、波型鉄製の新しい、またはよく手入れされている屋根かどうかを判別することで所得水準を推定している。同様に、Jain (2007) はインドの州都デラドゥンを対象に、IKONOS の高解像度画像を用いてプラスチック製の屋根を基準に貧困水準以下の家庭を判別し、貧困分布の推計を行っている。

2) インフラ投資の経済効果

途上国の物的インフラ投資・整備は経済発展実現のための重要な基盤であり、効果的なインフラ投資のために、それが経済成長・発展にどの程度貢献するか事前・事後の客観的で質の高い評価が不可欠である。近年、人工衛星の高解像度画像を活用して、インフラ投資の客観的な評価を行う試みがなされつつある。Duflo and Pande (2007) は、インドの灌漑・発電用ダムの建設に際して、ダムが貧困と農業生産に与える影響を、地理情報システム (GIS) を用いて地区単位のデータに基づく推計による検証を行っている。ダムの建設に

よる地区別の農業生産の変化と、所得や貧困状況など個人の経済厚生へ与える影響を詳細に推計した結果、インドにおけるダムの建設ではダム建設費用を正当化できるほどの経済的利益がないことを見出した。一方、Lipscomb, Mobarak and Barham (2013) は GIS を用いて、ブラジルの電化事業が経済発展を促す影響を持ち、電化による地価上昇を考慮すると利益が費用を上回ることを示した。Lipscomb, et al. の研究は一般均衡モデルを用いた従来の経済分析の枠組みに、人工衛星画像を含む GIS データを応用したものである。従来の経済学的な分析フレームワークに基づき、人工衛星画像を含む大規模でリアルタイムのデータによって拡張することで、インフラ投資の経済効果の評価がより広範囲かつ詳細に行える可能性を示している。

3) 災害、環境汚染、気候変動が経済活動に与える影響

2.1.2 節でも見たように自然災害による経済的な影響を正確に把握することは開発経済学の重要な課題である。さらに、環境汚染や気候変動についても、災害と同様に途上国では貧困リスクを高め長期的な成長阻害要因となる。これまで主に気象学や海洋学等の観測に用いられてきた自然現象の観測データを用いて災害の規模、環境汚染の程度や気候変動の推移を直接観測し、それらと既存の経済活動データを組み合わせることによって、正確かつ詳細な分析を行おうという試みがなされている。

Hsiang and Jina (2014) は、全世界を対象として約 60 年間の長期の国レベルのマクロデータと、人工衛星による気象データを組み合わせることで、熱帯低気圧直撃による被害の前後の経済成長率を比較した。人工衛星による気象データと地表、船舶、大気データを組み合わせ、位置と時間ごとの熱帯低気圧の被害を推計したデータを構築して用いている。分析の結果、復興時に経済成長がプラスになるという仮説は棄却され、途上国・先進国の両方で、国民所得の減少と災害前の成長のトレンドに回復するのに 20 年かかることを見出した。また、災害被害とマイクロレベルの行動との関係について分析した Guiteras, Jina and

Mobarak (2015) は、バングラデシュの洪水被害とその後の人々の行動について分析している。Guiteras et. al. は、NASA の MODIS の画像データを用いて、洪水のインパクトを推計しており、洪水被害のデータには、降雨量や自己申告の情報に代理変数として説明力が弱いことを検証している。これらの研究は、遠隔計測データを用いることで、従来の災害の経済効果の研究の多くとは異なる結果を得ている。特に豊富な自然観測データを組み合わせることによってより客観的な検証が行われることが期待される。

また、遠隔計測データで得られる多様な気候データは、既存のデータと組み合わせることで環境汚染や気候変動が経済厚生や経済活動に与える影響の分析にも用いられる。Kudamatsu, Persson and Stromberg (2012) は、人工衛星画像による気候データと人口保険調査を組み合わせ、アフリカ 28 か国を対象にして、気候変動が乳児死亡率に与える影響を推計している。その結果、乳児死亡率は母親の胎内にいる期間のマラリア流行、干ばつ、および出生時に飢饉に見舞われていることによって上昇することを見出した。出生前後の短い時間単位の環境データを使用することによって、乳児死亡率を高める要因をより詳しく検出している。また、Axbard (2016) は、インドネシアの漁師の所得機会と海賊行為の関係を、人工衛星画像データから構築した漁業条件指数を用いて検証した。Axbard は、漁師の漁場ごとの漁業条件を海洋学で用いられているクロロフィル濃度や海面温度（魚のエサの量や魚の成長に適した温度）を用いて作成して漁師の所得機会の代理変数として、各海域の海賊行為との相関を調べた。その結果、漁業による所得機会が減少すると海賊行為が増えることが見出された。気候変動と紛争や犯罪には相関があるとする先行研究と同じ結果となっているが、遠隔計測データを用いることによってこれまで説明が遅れていた、そのメカニズムを明らかにする試みを行っている。

3. 開発経済学における ビッグデータ活用の課題

CDR や遠隔計測データのようないわゆるビッ

グデータは、今後も開発経済学の強力なツールとして活用されていくと思われる。しかし、その試みは始まったばかりの段階であり、開発経済学への活用の際に直面する課題は多い。

まず、経済学全体に共通することであるが、従来の計量経済学的な分析手法をビッグデータに応用する際に生じる問題が挙げられる。例えば、人工衛星による地表の夜間の光量を経済活動の代理変数として回帰分析に用いる場合、個々のデータは地表を格子状に区切ったセルの明るさとなり、隣り合ったセルが影響を及ぼしあう可能性が生じる。その場合、データ間の相関（空間的自己相関）を考慮しなければ正確な推計ができない。地球全体のセル数は膨大な数になるので、短期間周期で記録されるデータを膨大な量のパネルデータとして分析に用いる場合に簡単な推計式でさえ困難になる。

また、ビッグデータは文字通りデータサイズが膨大であることから、実際に分析を行う際には、データ・クリーニングにも膨大な時間が必要である。それに関連して、ビッグデータには大きな計測誤差が生じている可能性があることに留意する必要がある。例えば、携帯電話の CDR の大きさを経済活動水準の代理変数とするような場合、地方に設置されたコールセンターの集中的な受信は誤差として排除する必要がある。また、人工衛星の画像データは、同じ地点のデータの時系列変化を使用したい場合に、気象条件が違えば正確なデータとして使うことができない。このように、ビッグデータを用いた分析には、データ処理の技術とデータそのものの問題点を慎重に考慮する必要がある。

最後に、経済学に限らない問題として、ビッグデータを扱う際の個人情報の保護に留意する必要がある。現在使用されているデータは匿名性が保たれた状態で公開されているが、例えば、人工衛星画像では個人の住宅の屋根の状態まで分かる高解像度のものがあるなど、プライバシー保護と、研究促進の線引きが難しいケースがある。個人情報・プライバシー保護と、ビッグデータの利用促進のバランスをどのようにとるのかはまだ試行錯誤の段階にある。

おわりに

経済学におけるビッグデータの活用はまだ始まったばかりであり、本稿で紹介した開発経済学におけるビッグデータの活用例はその試みの一部のみである。現在、開発経済学の幅広い分野で数多くの研究が行われており、研究の蓄積が進められている。また、機械学習による予測を取り入れた研究も急速に進んでいる。伝統的な経済学では理論モデルに基づいた仮説を立てデータを用いた検証を行い、その理論に基づいて予測を行う。一方、機械学習ではデータからパターンを読み解き、そのパターンから将来を予測する。すなわち、経済学と、統計学や計算機科学とは、経済現象のアプローチの仕方が全く異なっている。ゆえに、ビッグデータの経済学への応用は、単に使用できるデータが大幅に増えたというだけではなく、計算機科学のような他分野の成果を取り入れて経済学が発展していく一つの方向を示しているように思える。特に開発経済学は、途上国の直面する多くの差し迫った課題に取り組むため、多様な分析手法を取り入れながら発展してきた分野である。ビッグデータの応用によって開発経済学の発展がさらに進むことが期待される。

参考文献

- Ali, A., J. Qadir, R. Rasool, A. Sathiseelan, A. Zwitter and J. Crowcroft, 2016, Big data for development: applications and Techniques, Big Data Analytics ,2016 1: 2.
- Axbard, S., 2016, Income Opportunities and Sea Piracy in Indonesia: Evidence from Satellite Data, American Economic Journal: Applied Economics, Vol.8, Issue 2, pp.154-194.
- Bengtsson L., X. Lu, Thorson A, Garfield R, von Schreeb J., 2011, Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: a post-earthquake geospatial study in Haiti, PLoS Medicine, 8(8).
- Blumenstock, J. G. Cadamuro, and R. On, 2015, Predicting Poverty and Wealth from mobile Phone Metadata, Science, Vol.350, Issue 6264, pp.1073-1076.
- Blumenstock, J. 2016, Fighting Poverty with Data, Science, Vol.353, Issue 6301, pp.753-754.
- Chen, X., and W. Nordhaus, 2011, Using Luminosity Data as a Proxy for Economic Statistics, Proceedings of the National Academy of Science, Vol.108, Issue 2, pp.8589-8594.
- Donaldson, D. and A. Storeygard, 2016, The View from Above: Applications of Satellite Data in Economics, Journal of Economic Perspectives, Volume 30, Number 4, pp.171-198.
- Duflo, E., and R. Pande. 2007, Dams, Quarterly Journal of Economics, 122(2), pp.601-46.
- Eagle, N., M. Macy, and R. Claxton, 2010, Network Diversity and Economic Development, Science, Vol.328, Issue 5981, pp.1029-1031.
- Frias-Martinez, E., G. Williamson and V. Frias-Martinez, 2011, An agent-based model of epidemic spread using human mobility and social network information, the 3rd International Conference on Social Computing, SocialCom 2011, Boston, USA.
- GSMA, 2016, The Mobile Economy 2016, GSMA Report Global, The GSM Association (<https://www.gsmainelligence.com/research/?file=97928efe09cdba2864cdcf1ad1a2f58c&download>).
- Ghurye, J., Gautier Krings and Vanessa Frias-Martinez, 2016, A Framework to Model Human Behavior at Large Scale during Natural Disasters, 17th IEEE International Conference on Mobile Data Management.
- Guiteras, R. A. Jina, and A. Mushfiq Mobarak, 2015, Satellites, Self-reports, and Submersion: Exposure to Floods in Bangladesh, American Economic Review: Papers & Proceedings, 105(5), pp.232-236.
- Hamermesh, D. S., 2013, Six Decades of Top Economics Publishing: Who and How?, Journal of Economic Literature, 51(1), pp.162-172.
- Harari, M., 2016, Cities in Bad Shape: Urban Geometry in India, (<http://marroninstitute.nyu.edu/uploads/content/CityShapeHarariMarch2016.pdf>).
- Henderson, J. V., A. Storeygard and D. N. Weil, 2012, Measuring Economic Growth from Outer Space, American Economic Review, 102(2), pp.994-1028.
- Hilbert, M., and López, P., 2011. The world's technological capacity to store, communicate, and compute information. Science, 332, issue 6025, pp.60-65.
- Hsian, S. M., and A. S. Jina, 2014, The Causal Effect of Environmental Catastrophe on Long-Run Economic Growth: Evidence From 6,700 Cyclones, NBER Working Paper Series, No.20352.
- ITU, 2016, ICT Facts and Figures 2016, International Telecommunication Union, (<http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2016.pdf>)
- Jain, S., 2007, Use of IKONOS satellite data to identify in-

- formal settlements in Dehradun, India, *International Journal of Remote Sensing*, Vol.28, Issue 15, pp.3227-3233.
- Jean, N., M. Burke, M. Xie, W. M. Davis, D. B. Lobell, S. Ermon, 2016, Combining satellite imagery and machine learning to predict poverty, *Science*, Vol.353, Issue 6301, pp.790-794.
- Laney, D., 2001, 3 D Data Management : Controlling Data, Application Delivery Strategies, February 6, 2001, META Group.
- Kudamatsu, M., T. Persson, D. Stromberg. 2012, Weather and infant mortality in Africa, *CEPR Discussion Papers*, No.9222.
- Lipscomb, M., A. M. Mobarak, T. Barham, 2013, Development Effects of Electrification : Evidence from the Topographic Placement of Hydropower Plants in Brazil, *American Economic Journal : Applied Economics*, Vol.5, No.2, pp.200-231.
- Lu X, Bengtsson L., Holme P., 2012, Predictability of population displacement after the 2010 Haiti earthquake, *Proceedings of the National Academy of Sciences* Vol.108, Issue 29, pp.11472-11473.
- Marx, B., T. M. Stoker, and T. Suri. 2015, There is No Free House : Ethnic Patronage in a Kenyan Slum." [http : //www.mit.edu/~tavneet/Marx_Stoker_Suri.pdf](http://www.mit.edu/~tavneet/Marx_Stoker_Suri.pdf).
- Michalopoulos S., and E. Papaioannou, 2014, National Institutions and Subnational Development in Africa, *The Quarterly Journal of Economics*, 129(1), pp 151-213.
- Sundsøy, P., 2016, Can mobile usage predict illiteracy in a developing country? arXiv Preprint : 1607.01337.
- Pokhriyal, N., Wen Dong, Venugopal Govindaraju, 2015, Virtual Networks and Poverty Analysis in Senegal, Presented at NetMob 2015 for Data for Development (D4D) Challenge, arXiv : 1506.03401.
- Smith, C., A. Mashhadi, and L. Capra, 2013, Ubiquitous Sensing for Mapping Poverty in Developing Countries, Presented at NetMob 2013 for the D4D Challenge.
- Soto, V., V. Frias-Martinez, J. Virseda, E. Frias-Martinez, 2011, Prediction of Socioeconomic Levels Using Cell Phone Records, Vol.6787, pp.377-388, User Modeling, Adaption and Personalization, The Series Lecture Notes in Computer Science.
- The World Bank and DecondMuse, 2014, Big Data in Action for Development. ([http : //live.worldbank.org/sites/default/files/Big%20Data%20for%20Development%20Report_final%20version.pdf](http://live.worldbank.org/sites/default/files/Big%20Data%20for%20Development%20Report_final%20version.pdf))
- United Nations Global Pulse, 2013, Mobile Phone Network Data for Development.
- Varian, H. R., 2014, Big Data : New Tricks for Econometrics, *Journal of Economic Perspectives*, Vol.28, No.2, pp.3-28.
- Wesolowski, A., N. Eagle, A. J. Tatem, D. L. Smith, A. M. Noor, R. W. Snow, C. O. Buckee, 2012, Quantifying the Impact of Human Mobility on Malaria, *Science*, Vol. 338, Issue 6104, pp.267-270.
- 総務省 (2015) 『平成 27 年度 情報通信白書』