

■研究課題名

機械学習による新聞記事テキストを用いた企業動態情報大規模データベースの生成と分析

研究代表者：

鯉淵賢 (中央大学商学部・教授)

共同研究者：

飯塚修平 (Google)、**野中尚輝** (理化学研究所)、**星岳雄** (東京大学大学院経済学研究科・教授)

実施期間：2017年4月1日～2020年3月31日

【研究の概要】

従来の経済学・経営学などの研究分野では、研究者及び研究補助者が過去の一定期間の新聞記事を読み、そこから特定の企業や制度の変遷などの多面的な情報を抽出して総括する研究手法が事例分析として実施されてきた。しかし、人間が個別記事を読み下してまとめていくことから、必然的に、過誤に起因するノイズの発生や、人材面と予算面の厳しい制約に直面することが一般的であった。このため、研究対象の詳細な分析を可能にする事例研究とは、サンプル数が極めて少数に限定される傾向があった。本研究では、日本の上場企業の企業再編(企業再建、M&Aなど)の企業動態情報の分析を例にして、機械学習の手法により、従来の人間による事例研究の多くのプロセスを機械(プログラム)による判断に置き換えることにより、上場企業全社というような網羅的な研究対象の詳細な項目について、長期間にわたる事例分析を行うことを可能にしサンプル数の大幅な増加を容易にする2種類の機械学習プログラムを構築した。

第1の機械学習プログラムは、Hoshi, Koibuchi, and Schaede(2011, 2018)において1980年から2010年の日本経済新聞の「再建」という語句を含む記事テキスト(約16万記事)から人間による作業によって上場企業の企業再建に関するものとして抽出された記事をサンプルとして、2011年から2016年の記事テキストから上場企業の企業再建記事を生成するウェブ上のAPIを構築した。これは、記事テキストを読み込ませると、その記事が「企業再建」に関する記事であるかどうかの確率を表示する。研究者は自身で確率の閾値を設定して、記事をサンプルとすることが可能である。

この第1の機械学習プログラムは、人間による作業によって抽出しラベル付けされた十分に多数の記事サンプルが入手可能であることを前提としていた。また、このため、「企業再建」という特定のテーマのみについて判別が可能となっていた。これらの点について、①研究者自身が比較的少数のラベル付けデータを作成することによって、②任意のテーマについて記事抽出を可能としたのが、第2の機械学習プログラムである。この第2の機械学習プログラムは、「M&A(企業の合併買収)」をテーマに、「合併」「買収」「営業譲渡」「資本参加」「出資拡大」という

M&Aを表す5つの語句のいずれかを含む1980年から2016年までの記事を抽出し、各年で数百程度の記事を研究者が日本の上場企業のM&Aに関する記事であるかを判断してラベル付けデータを作成すると、それをサンプルとして、残りの新聞記事がM&Aに関する記事かどうかの確率を表示して、任意の閾値に基づいてM&Aに関する記事サンプルを生成することが可能となった。

本研究により得られる主な成果は次の2つであると予測される。第1に、新聞記事から、企業の動的な情報を多面的に抽出して分析しまとめる作業を、機械学習に基づく新たな手法に生成する手法を確立したことである。これにより、従来手作業で行われてきた事例研究のサンプル数を劇的に増大させる可能性があり、研究者が与えた特定のテーマに対して、可能な限り全事例を抽出して分析対象とすることができるようになり、サンプルバイアスの低減を期待できる。また、事例研究によって得られた特徴の全サンプルにおける割合などの量的な情報を得ることができるようになる。第2に、日本の上場企業の企業再建やM&Aを含めた企業再編の変遷について、より精緻なデータベースを構築し、分析することが可能となる。これは企業金融分野や企業経済学の学術分野における追加的な貢献となる。

■研究課題名

日米における商業不動産価格のダイナミクスの比較研究

研究代表者：

清水千弘（日本大学スポーツ科学部・教授）

共同研究者：

西村清彦（政策研究大学院大学・特別教授）、渡辺努（東京大学大学院経済学研究科・教授）

実施期間：2015年4月1日～2020年3月31日

【研究の概要】

不動産価格の大規模な上昇と下落は、多くの主要国の経済運営に対して深刻な影響をもたらしてきた。わが国における1980年代中ごろから始まった不動産バブルは、20世紀最大のバブルと言われた。そして、その後においては、「失われた10年 (lost decade)」と揶揄されたように、長期的な経済の停滞に直面した。このような問題は、1990年代のスウェーデンの経済危機や21世紀に入ってからの米国を中心とした不動産バブルの生成と崩壊によってもたらされた世界的な金融危機と経済停滞など、多くの国が共通に経験したことであった。しかし、いずれの国においても、不動産価格がいつからどの程度上昇し、そして、どの程度下落したのかといったことを「正確に」把握することができなかった。ましてや、国際的な政策連携の重要性が高まる中で、その資産価格変動を比較することができない状況に置かれてきた。そのような中で、2009年のG20において、公的部門が不動産価格指数を推計し、公表していくことが決定されたことを受けて、国際通貨基金や国際決済銀行などの国際機関が中心となり、国際的に比較可能な不動産価格指数の整備が進められてきた。本研究プロジェクトは、これらの国際機関とも共同し、日本・米国の実際のマイクロデータを用いた不動産価格指数の推計とダイナミクスの解明を進めたものである。

不動産価格指数の整備に対する期待は、かねてから大きかった。しかし、不動産価格は、経済測定分野において最も測定が困難な対象であること、既存の公的統計との連動性も大きいことから、実際の整備まで至ることがなかった。また、実際の不動産価格指数の推計においては、国ごとに不動産市場の流動性の程度やストックの構成、フロー（建設量）の規模や性質、そして、情報整備の状況やその入手のコストなどが異なるために、配慮すべき要因も多様であった。そのために国際的に足並みをそろえて進めることの技術的な問題が、とりわけ大きかったともいえよう。そのような市場の異質性を前提としつつも、金融政策、マクロ経済政策の国際的協調の重要性が増す中で、相互の市場を比較観察可能な価格統計を整備しようとすることの重要性は上昇してきていたのである。その中で、不動産価格指数の国際的な整備は、一連の

国際機関での議論を通じて各国の共通の課題とそれぞれの国の特有の問題を整理し、それぞれが抱える政策的な課題に対応可能な情報資源を配信していくことで、政策当局と市場とのギャップを改善しようとしたものであった。とりわけ、米国のサブプライム問題に端を発した世界同時不況に見舞われる中で、この問題への関心は一気に高まることとなった。

不動産価格指数を経済統計として整備しようとした場合には、統計としての正確性と整備費用を含めた効率性といった二つのトレードオフ問題に直面する。それでは、不動産価格指数にはどのような正確性が要求されるのであろうか。どのような手法が適用可能なのであろうか。本研究プロジェクトでは、日本と米国およびそれ以外の主要国にも注目し、入手可能な不動産価格に関する情報資源とその課題、さらにはその限界を踏まえつつ、実際にデータを用いて新しい推計技術の開発を行った。まず新しいデータ資源としては、日本・米国・欧州・アジアで成長するREIT(不動産投資信託)市場において取得可能なデータを用いた価格指数の推計可能性と新しい推計モデルの提案をおこなった。また、推計手法としては、日本や欧州ではヘドニック法を用いた価格指数の開発がすすめられたが、米国ではリピートセールス価格法が採用されていた。リピートセールス価格法を適用する際の推計上の課題としては、「経年減価バイアス」が挙げられる。そこで、このような問題に対応可能な新しいリピートセールス価格法の推計方法を開発した。さらに、小地域単位での不動産価格指数の推計手法の開発を行った。多くの不動産価格指数が東京・大阪といった都市単位での指数として公表されているが、1kmメッシュなどの小地域レベルでの価格指数を推計するための技術を開発した。また、不動産価格は、市場では土地と建物一体で取引された価格しか観測できない。しかし、価格変動の多くの部分は土地から発生しており、不動産の経年減価は建物にしか発生しない。そのため、土地と建物の価格を分解していくことが必要となる。そこで、土地・建物から構成される不動産価格を、土地と建物に分離することが可能な新しいヘドニック価格指数の推計方法を、戸建て住宅市場、マンション市場、オフィス市場を対象として開発した。

以上の一連の分析は、15本の学術論文として公表されるとともに、一連の研究成果の集大成としてSpringerから下記の学術書として2020年1月に公刊した。

Diewert, W.E., Nishimura, K.G., Shimizu, C., Watanabe, T., (2020), *Property Price Index*, Springer.