

統計分析用特許データベースの進展

IIP パテントデータベース

東京大学工学系研究科教授

元橋 一之

PROFILE

東京大学工学修士、コーネル大学MBA、慶応大学博士（商学）。経済産業省、OECDエコノミスト、一橋大学助教授などを経て2006年から現職。主な著書に「11 イノベーションの実証分析」（東洋経済新報社）、「日本経済競争力の構想」（日本経済新聞社）など。



1 はじめに

特許データベースは企業などにおける先行技術調査などで利用されることが多い。また、最近では特定技術分野におけるパテントマップの作成やそれらの情報を利用した技術経営戦略の立案などに用いられてきている。このような特許データに対するニーズを反映して、一部の企業や特許を専門とする調査会社では、独自のデータベースを構築して、社内での活用や外部ユーザーに対する提供を行っている。

特許データベースに対するニーズはこのような企業ユーザーのみならず、大学などにおける社会科学研究者の間においても高まっている。特許データは企業等における詳細な研究開発活動を示すデータとして、イノベーションに関する経済学的な分析を行う研究者にとって貴重な情報源である。また、特許データを用いることによって、プロパテント政策とイノベーションの関係など、政策分析を行う上でも有益なツールである。

「IIP パテントデータベース」は、こうした要望に応えるべく、後藤晃氏（東京大学名誉教授）と筆者が中心となって構築した、我が国で最初の公開型の本格的な実証分析用特許データベースである（当データベースは、財団法人知的財産研究所のHP上で公開されている。構築プロセスの詳細は、Goto and Motohashi(2007) “Construction of a Japanese Patent Database and a first look at Japanese patenting activities”, Research Policy, Volume

36, Issue 9, November 2007, Pages 1431-1442 を参照されたい）。

2 データベースの概要とアップデート

公開当初、IIP パテントデータベースに含まれる特許データは、1964年1月以降の出願から2004年1月時点で出願公開された約902万件のデータであった。

そこには、特許庁の整理標準化データから抽出された、特許出願データ（出願番号、出願日、審査請求日、技術分野、請求項数等）、特許登録データ（登録番号、権利消滅日等）、出願人データ（出願人名、個法官コード、国・県コード等）、権利者データ（権利者名等）、引用情報データ（引用・被引用特許番号等）が含まれている。

整理標準化データはおおむね月2回のペースで公表されていることから、当データベースについても適宜アップデートを行ってきている。2010年8月時点においては、2008年8月公開特許のもの（整理標準化データ平成20年度第12回提供分）までデータの更新を行っているところである。また、発明者データの追加などデータベースの拡充についても行っている。

ただし、このアップデート・ファイルは、現時点では完全版ではなく、（株）人工生命研究所のHP上で公開され、そこに設置された掲示板にてユーザーからのバグ出しや問題提起等の議論を重ねることで、質の改善が

図られている。そこでの議論は、定期的開催される実証研究の専門家からなる運営委員会できりまとめが行われ、今後の改善に役立てられる。

3 IIP パテント データベースの拡充

東京大学においては、経済産業省の委託研究（産業技術調査事業）などを受けながら、上記の IIP パテントデータベースの更新・拡充とともに他の統計調査との接続などのデータベース基盤整備事業に取り組んでいる。前述したようにこのデータベース基盤の中核的なデータとして、IIP パテントデータベースをはじめとして、科学技術研究調査（総務省）と企業活動基本調査（経済産業省）の企業レベル個票データの接続を行っている。

まず、科学技術研究調査の企業その他、大学や公的研究機関における研究開発活動を総合的に調査しているものであるが、ここでは資本金 1 億円以上の企業に対して行われている企業等 A の調査項目について、1984 年からのパネルデータを作成している。科学技術研究調査においては、科学コードという番号で企業データの整理が行われているが、コードの付け替えが行われている年があり、パネルデータの作成にあたっては新旧コードの対応関係を把握することが適当である。この作業によって、5414 企業のアンバラストパネルデータ（最多年の 2002 年の企業数が 3650）を作成した。

一方、企業活動基本調査は 1991 年に開始された比較的新しい統計調査であるが、こちらは永久企業番号という期間を通じて統一したコード体系が整備されており、パネルデータの作成は容易である。資本金 3000 万円以上でかつ従業員数 50 人以上の製造業または卸小売業（2001 年から一部のサービス産業に対して業種が拡大）に属するすべての企業に対する調査であり、毎年約 2.5 万社のサンプル数となっている。

これらの統計調査の企業パネルデータに IIP パテントデータの出願人（日本に所在する企業のみ）約 60 万社を接続させたものがイノベーションデータベース基盤の

中核的な構成要素となっている。

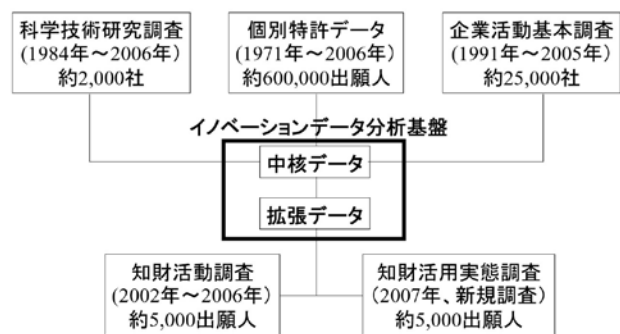


図 イノベーションデータ分析基盤のフレームワーク

また、これらまでのデータ接続作業としては、これらの中核データの整備に加えて、「知的財産活動調査」（特許庁）や新規に行ったライセンスに関するアンケート調査（知財活用実態調査）を接続し、データ拡充を行っている。知財活動調査は特許庁において行われている承認統計であり、毎年一定数の特許、実用新案、商標及び意匠の出願を行っている出願人（企業その他、大学や個人発明家含む）に対してして毎年行われている。内容は、各種知財の出願状況やライセンスなど知財利用状況、知的財産活動費や知財侵害の実態に関するものである。このデータは特許庁における出願人番号をベースとして管理されているため、中核データの個別特許データ（同じ出願人番号の情報が存在する）との接続が可能である。

知的財産活動調査は出願数やライセンスに用いられる特許数など定量的な情報として有益なものが存在するが、特許データをイノベーション分析に用いるためには、特許性向（すべての発明が特許化されるわけではないので、発明案件のうちどの程度の割合のものを特許出願するか）や特許利用の実態に関する詳細な情報（例えば特許保有の目的として防御的な目的なのかライセンスアウトを目的としているのか）が必要となる。これらの情報は知的財産活動調査から得ることができないので、新規にアンケート調査（ライセンス活動実態調査）を行いデータベースの拡充を行った。



4

イノベーションデータベースを用いた分析事例

それではこのようなデータを活用することによって、どのような分析が可能になるのでしょうか？まず、イノベーション分析に関する基礎的な集計データとして、産業・技術コンコダンステーブルをあげることができる。特許データはIPC分類などの技術分類に従って集計することは可能であるが、それがその発明がどの産業によって行われ、どの産業によって使われているのかという点は不明である。この技術分類と産業分類の対応関係は、イノベーションの分析を行っていく上で重要な情報となる。

特許データを用いたコンコダンステーブルは、OECD Technology Concordance や Yale Technology Concordance などが存在するが、これらはカナダ特許庁のデータを用いている。カナダ特許庁は1972年から1995年までの間、特許審査の過程においてIPCコードの格付けとともに当該特許の製造産業（Industry of Manufacture: IOM）と当該特許のユーザーセクター（Sector of User: SOU）の特定を行った。このような作業の結果、約30万件の特許についてIPCコードと産業分類（SICコード）の対応関係に関する情報が存在し、例えばこのデータを用いてIOMとSOUの関係をISICベースでマトリックスにしたものがOECD Technology Concordance (OTC) である。

ここでは、OTCのようにカナダ特許庁のデータを用いるのではなく、企業レベルの特許保有と産業別売上高のデータを用いてIPCコードと産業分類のコンコダンステーブルの作成を試みた。本来は、カナダ統計局のデータのように特許毎のIOMやSOUが分かれば理想的であるが、日本にはそのようなデータが存在しない。従って、企業レベルの集計データを用いてIPCと産業分類の関係について明らかにしたものである。

具体的にはIPC分類のサブクラスレベルと企業活動基本調査の3桁産業分類のテーブルを1995年と2005年のデータで作成した結果、おもに電気機械工

業と化学工業の分野でイノベーションが盛んにおこなわれるようになってきていることが分かった。また、コンコダンス表の作成にあたっては、技術分野に対応する産業が当該技術のサプライヤー（IOM: Industry of Manufacturer）なのか、技術のユーザー（SOU: Sector of User）なのかといった情報が重要である。これは図のデータベース構造の中で知財活動調査が加わったことによって、企業毎に所有特許の活用状況（自社実施、インライセンスやアウトライセンスの状況）が分かるようになっており、この情報を活用することによって、より詳細なテーブルを作成することができる。日本企業の特許利用率（保有特許に占める自社実施かライセンスによって利用している特許の割合）は平均6割といわれているので、特にSOUの分析については当該データを用いて補正をおこなうことが重要である。

また、研究開発税制や補助金の効果分析などの各種技術政策に関する評価分析や企業における知的財産マネジメントに関する定量分析など多様なイノベーションにかかる分析テーマに取り組むことが可能となる。これらの政策担当者と企業の実務家の双方によって有益な分析事例の紹介も行っているところである。

5

課題と今後の取り組み

IIPパテントデータベースは実証分析用のデータベースとして極めて有用であるが、利用に際しては次のような問題を含んでいることも理解しておかなければならない。

まず、出願人名に関して、同一企業にもかかわらず、出願人名や住所の表記に揺れがある場合や異なる事業所から出願されている場合などに、異なる出願人番号が付けられていることがある。IIPパテントデータベースの構築に当たっては、特許庁の出願人コードの不備について、特に大企業についてはかなりの精度で名寄せを行っているが、完全と呼べるまでには至っていない。

欧米の特許データベースも同様な問題を抱えており、

各国において効率的な名寄せ手法に関する研究が進められている（その成果の一部は、OECD や EPO が毎年主催している国際学会等において報告されている）。

他にも、IIP データベースの出願人名は、社名変更があった場合、最新の出願人名・番号に書き換えられているため、出願時の出願人が識別できないケースが生じることもある。

今後は、こうした問題についても取り組んでいくとともに、ユーザー間での活発な議論を通じて、データベースの質を高めていく予定である。なお、現在公開中のβ版はバグ出しが一通り終了した後、知的財産研究所のHP上で運営されることになっている。

また、特許データと企業データの接続については、これまで企業活動基本調査の対象企業に限られていた。従って、今後は日本の存在するすべての事業所・企業を包含する事業所・企業統計との接続について行っていくことを予定している。これは OECD において進められている特許データとビジネスレジスターの接続とデータ分析プロジェクトにも対応するものであり、我が国企業の特許活動の状況を欧米における状況と比較できるという点でも意義が高い。

更に、OECD においては、商標に関するデータベースの構築など、特許以外の知的財産権データに対する取組も始まっている。国際的に特許を中心とした研究者向けのイノベーションデータベースの整備が進む中、我々としても積極的に新しいプロジェクトに取り組み、その成果を公表していきたいと考えている。

