

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『IT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。



財団法人知的財産研究所

山内 勇

PROFILE

2006年より現職。特許統計を用いた調査研究に従事。



1 はじめに

従来、特許データは先行技術調査や権利調査等の目的で利用されることが多かった。それが近年では、自社や他社の技術動向を分析し、研究開発戦略や特許出願・取得戦略の意思決定に関する重要な判断材料としても用いられるようになってきている。

一部の大企業や特許を専門とする調査会社では、社内に独自の特許データベースを保有し、知財戦略等の作成に役立てている。

企業・経済の成長にとってイノベーションの重要性が認識されてきた今日では、この豊富な技術情報を含む特許データを用い、さらに詳細かつ厳密な分析の上で企業戦略・政策決定に関する議論を行うことが望まれる。

「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)や元橋一之氏(東京大学教授)ら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。

当データベースは、2008年5月にアップデートが行われた。これにより、収録期間が2007年5月公開のもの(整理標準化データ平成19年度第5回提供分)へと延長されるとともに、発明者データが追加された。

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を用いた研究

特許データは、そのみでも、企業別・技術分野別の

特許出願件数、審査請求率、登録率、権利維持期間等の動向を把握したり、特許集中度などの指標を用いた分析を行ったりすることが可能である。また、パテントマップの作成など、ある程度高度な分析をすることもできる。しかし、他のデータベースと接続することで、より多様な質の高い研究が可能となる。

例えば、基礎的な企業情報を含む財務データと接続すれば、企業属性（産業、規模、利益率、負債比率など）と特許活動との関係を分析することができる。特に、研究開発データと併せて用いることで、イノベーション活動のインプットとアウトプット両面を考慮した分析が可能となる。

こうした特許データを用いたイノベーション研究は、特に欧米で活発に行われてきた。米国のNBERパテントデータベースや欧州のPATSTATデータベースなどを利用し、イノベーションの決定要因やプロパテント政策の評価など様々な研究がなされてきた。

近年では、我が国のイノベーション活動を分析する目的で、IIPパテントデータベースを利用した研究も増えてきており、今後の更なる発展が期待されることである。また、上述のように、そのためのデータベース自体の整備も進展してきている。

今回のIIPパテントデータベースのアップデートで最も大きな変更は、発明者ファイルが追加されたことである。

発明者はイノベーションの最も根源的な要素であり、発明者レベルの分析を行うことはイノベーションの発生メカニズムを理解するうえで非常に重要である。企業レベルの分析では、発明者間の違いをすべて無視することになってしまう。

例えば、同じ内容の報奨制度を導入しても、所属している発明者の過去の発明経験や分野等が違えば、企業間でもその効果に差が出てくるはずである。

また、共同研究の効率性や発明者間のネットワーク性に関する分析などにも、発明者データが役に立つ。

こうした研究に際して、引用データを併せて用いることで、当該特許の重要性や知識の流れなども知ることができる。すなわち、引用した特許が多ければそれだけ多くの技術を基に生まれた高度な発明と考えることもできるし、ある特許がどの発明から生じたか、どの程度の速度（タイムラグ）で生じたかといったことまで分かる。

また、引用される数が多ければそれだけ重要な特許と考えることもできる。

4 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。

最後に、国内の研究グループの中には、例えば、テキストマイニング等により、明細書から発明者の引用情報を抽出するといった研究を行っているグループもあり、我が国のイノベーション研究の一層の進展のためには、こうした研究グループとの連携を図っていくことも必要であると考えられる。また、東京大学においては「イノベーションデータベース基盤の整備」事業が進んでおり、IIPパテントデータの出願人情報と政府統計や公開財務諸表とのリンケージ作業が進んでいる。IIPパテントデータの活用によって、イノベーションプロセスの研究が一層進むことを期待したい。