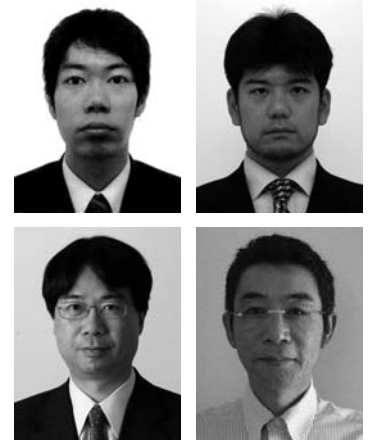


特許成立性に寄与する客観的指標の実証分析



田中 克幸[※] 田中耕一郎^{※※}
 会員 清水 初志 加納 信吾^{※※※}

1. 序論

これまで、特許成立性に関する議論は、新規性・進歩性の特許要件や記載要件等について、判例研究等のケーススタディによる分析が中心であった⁽¹⁾。

一方、特許成立性に関して、明細書等から抽出した数値化情報（以下、「特許情報」という。）を用いて分析を行った研究は、安彦ら⁽²⁾、永田ら⁽³⁾など、ごく少数に限られる。ここで、特許情報は、審査経過の各ステージにおいて補正等により変化するが、これら二つの先行研究は図1に示すように出願時や訴訟時における特許情報を対象としており、特許査定又は拒絶査定が下される時点（以下、「査定時」という。）における特許情報を用いた分析は、これまでなされていなかった。これは、図1に実線（明細書等が公開される）と点線（明細書等が公開されない）にわけて示すように、拒絶査定が下された案件の明細書等（A'で示す）は公報として公開されないため、**査定時**における特許査定と拒絶査定に関する特許情報を比較することが困難であったことが大きな要因であったためと考えられる。さらに、特許成立性に関して技術分野に関係なく一律

に分析することは果たして適切なのかと疑念が生じるが、これら二つの研究は、技術分野別の比較については触れていない。

また、個別の議論を離れ網羅的分析の手法により、特許成立性の観点から技術分野の特徴を議論した研究は、金井ら⁽⁴⁾などごく少数の研究が存在する。しかし、金井らは技術分野の組み合わせから特許成立性を議論したものであり、技術分野別の特許成立性の比較に際し、特許情報による評価を行った研究は従来なかった。

以上をまとめると、

- 1：**査定時**における特許情報を用いた特許成立性の分析
- 2：**特許情報という手段**からの特許成立性に関する技術分野の比較分析

に関する研究は、従来行われていなかった。

そこで、本研究においては、**査定時**における特許情報等の客観的に入手できる数値化情報を用いて、技術分野別に特許成立性に影響をあたえる要因及びその寄与度を分析することを目的とした。

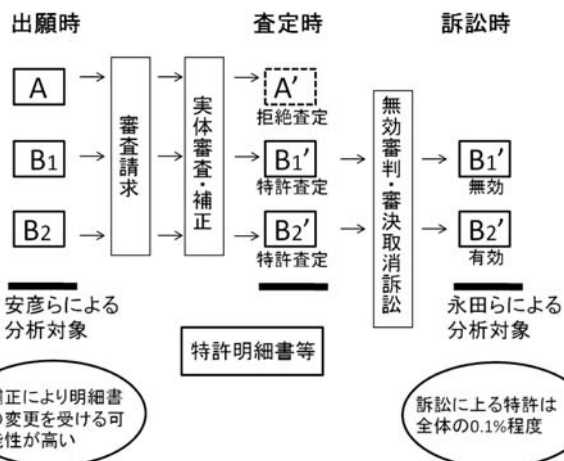


図1 先行文献の分析対象時点

2. 研究手法

本研究においては、上記目的に対する研究手法として、**査定時**における客観的な定量的指標を説明変数として採用したロジスティック回帰分析を行った。

この研究手法をとった理由は、ロジスティック回帰

※ 元東京大学大学院新領域創成科学研究科
 ※※ 特許庁審判部審判第22部門（生命工学）審判官（平成22年3月31日まで東京大学大学院新領域創成科学研究科准教授）
 ※※※ 東京大学大学院新領域創成科学研究科特任教授
 なお本稿は、東京大学大学院新領域創成科学研究科メディカルゲノム専攻 バイオ知財コースにおいてなされた研究を基に執筆されたものである。

モデルの構築により、回帰モデルに採用した説明変数および回帰係数の値の大きさから、特許成立性に影響を与える要因及びその寄与度を分析することが可能といえるからである。ここで、回帰モデルのうちロジスティック回帰モデルを用いたのは、従属変数に、特許査定を受ける・拒絶査定を受ける、の2値の質的変数を設定したためである。

2.1 回帰モデルにおける説明変数の採用

回帰モデルの構築のためには、特許成立性との関係を議論できるような適切な説明変数を採用する必要がある。そこで本研究においては、実際の特許審査において問題となる新規性・進歩性の特許要件、記載要件を考慮した上で、特許成立性に影響を与える要因として「請求項記載様式」、「出願人属性」、「先行技術文献数」、「特許庁との対応」に着目し、各要因から客観性を持って数値化情報が得られるものとして表1に挙げる説明変数を採用した。

表1 本研究で採用する説明変数一覧

請求項記載様式を表す指標 【請求項1】文字数×0.01 全請求項平均文字数×0.01 請求項数	【請求項1】文字数に100を除いた値 全請求項平均文字数に100を除いた値 請求項数
出願人属性を表す指標 出願人ダミー変数	大学等の非営利企業を1、営利企業を0としたダミー変数
先行技術を表す指標 引用特許文献数 引用非特許文献数	特許・拒絶査定時に審査官によって参照された特許文献数 特許・拒絶査定時に審査官によって参照された非特許文献数
特許庁との対応を表す指標 補正書提出回数 意見書提出回数	拒絶理由通知が出される後の、出願人による補正書提出回数 拒絶理由通知が出される後の、出願人による意見書提出回数

2.2 推計モデル

上記の各説明変数に対して、ロジスティック回帰分析を行う前に、単回帰分析を行ったところ、いずれも有意になった。しかし、説明変数間の相関行列を確認したところ、「【請求項1】文字数×0.01」と「請求項平均文字数×0.01」との間で、また「意見書提出回数」と「補正書提出回数」との間で強い相関が生じていた。互いに強い相関を示す変数を同時に回帰式の説明変数として採用すると、多重共線性の問題⁽⁵⁾から回帰式の推定結果が不安定となる。このため、「【請求項1】文字数×0.01」と「請求項平均文字数×0.01」のうち一方を選択、かつ「意見書提出回数」と「補正書提出回数」のうち一方を選択するような4種類のモデルを構築したところ、いずれも大差が認められなかった。

そこで、本論文においては4種類のモデルの代表として下記で表わされる回帰モデル式を用いた結果を示

す。ここで、特許査定を受ける確率を p ($0 \leq p \leq 1$)、拒絶査定を受ける確率を $1 - p$ ($0 \leq p \leq 1$) とすると、予測確率 p は $p = 0.5$ を境に、 $p > 0.5$ で従属変数1 (= 特許査定) に分類され、 $p < 0.5$ で従属変数0 (= 拒絶査定) に分類される。これにより予測値の実測値に対しての正答率を求めることができる。

$$\begin{aligned} \text{Log}(p/(1-p)) = & a + \beta_1 \times \text{【請求項1】文字数} \times 0.01 \\ & + \beta_2 \times \text{請求項数} \\ & + \beta_3 \times \text{出願人ダミー変数} \\ & + \beta_4 \times \text{引用特許文献数} \\ & + \beta_5 \times \text{引用非特許文献数} \\ & + \beta_6 \times \text{補正書提出回数} \end{aligned}$$

2.3 データセット

本特許で用いるデータベースは、特許庁電子図書館「IPDL」、野村総合研究所「NRI サイバーパテント」である。

まず、NRI サイバーパテントを用いて、文献種別、出願時期、技術分野、査定種別についての各種条件設定から分析対象特許群を抽出した。次いで、IPDLを用いて、上記で抽出した分析対象特許群について、「【請求項1】文字数」などの説明変数に対するサンプルデータを抽出した。

2.3.1 分析対象特許群の抽出

NRI サイバーパテントの特許データベースから、基本となる分析対象特許群の抽出を2008年12月1日に行った。

ここでは、国内出願、国際出願を問わず、2002年3月に日本国特許庁に出願された全ての特許を抽出の対象とした。技術分野については、①機械分野(筆頭IPC:F01)、②IT分野(筆頭IPC:G06F17/+G06F19/)、③医療機器分野(筆頭IPC:A61B)、④医薬化粧品分野(筆頭IPC:A61K)、⑤ニューバイオ分野(筆頭IPC:C12N+C12P+C12Q かつ IPC:C12N15/を含む)、⑥オールドバイオ分野(筆頭IPC:C12N+C12P+C12Q かつ IPC:C12N15/を含まない)、査定種別については、①特許査定、②拒絶査定を受けたものをそれぞれ抽出した。これにより、査定種別および技術分野別に、以下の表2で示される分析対象特許群のデータセットを得た。

なお、本研究では、技術分野の条件設定で「筆頭

IPC」を主に用いた。各特許には通常、複数の「IPC」が付与されており、「IPC」の条件設定を用いるなら、複数の技術分野から重複してカウントされるおそれがあるためである。また、本研究においてバイオ系分野では、筆頭 IPC により C12P+C12Q+C12N と定義したが、これには遺伝子工学を含む技術としてのいわゆる「ニューバイオ分野」と含まない技術としてのいわゆる「オールドバイオ分野」の双方を含む。これらは同じ筆頭 IPC の条件設定で定義されるものの、技術的特徴を異にすると考えられる。そこで、バイオ系分野の中で、IPC が C12N/15（突然変異または遺伝子工学；遺伝子工学に関する DNA または RNA、ベクター、例、プラスミド、またはその分離、製造または精製；そのための宿主の使用）を含むものを「ニューバイオ分野」、含まないものを「オールドバイオ分野」とした。

特許査定、拒絶査定を受けた分析対象特許群は、それぞれ、NRI サイバーパテントの「複合検索」の画面において、「査定種別」で「登録」または「拒絶」を指定することで抽出した。しかし、これにより抽出された分析対象特許群は、特許無効審判や拒絶査定不服審判、審決取消訴訟によって、NRI サイバーパテントのデータベース上のものから査定種別が変化している場合がある。この場合、査定種別は最新のものを用いた。すなわち、拒絶査定を受けた後、拒絶査定不服審判により特許が認められたものは特許査定に、特許査定を受けた後に特許無効審判、審決取消訴訟によって拒絶となったものは拒絶査定に分類した。

表 2 抽出された分析対象特許群

	特許査定件数	拒絶査定件数	計
機械分野	124	84	208
IT 分野	113	200	313
医療機器分野	158	76	234
医薬・化粧品分野	107	121	228
ニューバイオ分野	59	64	123
オールドバイオ分野	34	32	66
全体（計）	595	585	1180

2.3.2 分析対象特許群からのサンプルデータの抽出

次いで、以上によって入手した分析対象特許群の一つずつに対し、IPDL の「特許・実用新案 DB」を使用して、各説明変数に対するサンプルデータを抽出した。

しかし、ここでの問題として、図 1 で示されるように、特許査定を受けた分析対象特許群は、査定時における特許明細書等 B' が「特許公報」として公開される一方で、拒絶査定を受けた分析対象特許群は、査定時における特許明細書等 A' が公報として公開されない。

そこで、拒絶査定を受けた特許明細書等については、査定時における特許明細書等 A' についてのデータを新規に取得する必要がある。これは、IPDL の「特許・実用新案特許査定 DB」を使用して基本となる出願時における特許明細書等 A を参照し、次いで IPDL の「審査情報照会」によって手続補正書を参照し、出願時の特許明細書等 A について、補正部分を最新の補正内容により置換することで特許明細書等 A' についてのデータを取得することにより行った。ただし、IPDL の「審査情報照会」により参照できる手続補正書は 2003 年 7 月以降に提出されたものであり、最新の手続補正書がそれ以前に提出された場合には、拒絶査定を受けた時点における特許明細書のデータは取得できない。そのため、拒絶査定を受けたものについて、一部データは欠損する⁽⁶⁾。

3. 結果

先の「2.2 推計モデル」でモデルに基づいて、各説明変数を強制投入したロジスティック回帰分析を行った。ここでは、本研究で分析対象とした全ての技術分野である「全体」及び個別の分野である「機械分野」、「IT 分野」、「医療機器分野」、「医薬化粧品分野」、「ニューバイオ分野」、「オールドバイオ分野」について分析を行った。

表 3 は、各技術分野のロジスティック回帰分析の結果をオッズ比で表したものである。ここで、オッズ比とは、自然対数をとったものが回帰式における説明変数の係数に相当し、各説明変数の特許成立性に対する寄与度の大きさを表す。例えば、表 3 の「全体」の結果によると、「【請求項 1】文字数×0.01」のオッズ比は 1.143 であることから、「【請求項 1】文字数×0.01」が一単位増えると、すなわち【請求項 1】文字数が 100 文字増えると、特許成立性は 14.3% 高まることがわかる。また、「出願人ダミー変数」のオッズ比は 1.723 であることから、「出願人ダミー変数」が 1 であると、すなわち大学等の非営利企業の発明であると、特許成立性は 72.3% 高まることがわかる。

表3 ロジスティック回帰分析の結果（オッズ比）

	全体	機械	IT	医療 機器	医薬 化粧品	ニュー バイオ	オールド バイオ
【請求項1】文字数×0.01	1.143***	1.799***	1.270***	1.219**			
請求項数	0.968***		0.971	0.975	0.929***		
出願人ダミー変数	1.723***		2.420		0.304	2.370***	
引用特許文献数	1.101***	1.258***	1.285***	1.145***		1.341***	
引用非特許文献数						1.127	
補正書提出回数	3.169***	2.738***	2.549***	3.457***	4.522***	5.197***	8.347***
NagelkerkeR ²	0.286	0.408	0.311	0.300	0.393	0.454	0.467
Hosmer&lemeshow	0.786	0.244	0.083	0.210	0.106	0.081	0.437
正答率	70.2	72.5	71.7	75.2	74.0	77.2	73.4

注：***は1%有意水準，**は5%有意水準，*は10%有意水準，無印は20%有意水準を表す。

ここで、いずれの技術分野の分析においても、モデル χ^2 検定の結果は $p < 0.01$ で有意であり、作成された回帰式の有意性が保証された⁽⁷⁾。また、いずれの技術分野の分析においても Hosmer&lemeshow の検定結果⁽⁸⁾が $p > 0.05$ を超え良好であり予測精度が高いことが示された。また、いずれの技術分野の分析においても、実測値に対して予測値が±3SDを超えるような、はずれ値は存在しなかった。

今回得られた結果を基に、興味深いものとして以下のことがわかった。

- ① NagelkerkeR²の値⁽⁹⁾は、「全体」より個別の技術分野の方が高かった。すなわち、特許情報等を用いて特許成立性を分析する際、個別の技術分野別に分析した場合の方が、モデルの説明力が良い。
- ② 機械分野やIT分野においては他分野に比べて「【請求項1】文字数×0.01」⁽¹⁰⁾の特許成立性への寄与度が大きい一方、医薬化粧品分野やバイオ分野などのライフサイエンスの分野においては他分野に比べて「補正書提出回数」⁽¹¹⁾の寄与度が大きい。
- ③ 一般的に、引用特許文献数が増えると特許成立性が高まる。
- ④ ニューバイオ分野では、出願人が大学等の非営利企業であると特許成立性が高まる。

4. 結論

本研究においては、査定時における客観的な定量的指標に基づいて、技術分野別に特許成立性に影響を与える要因及びその寄与度が異なることを初めて明らかにした。すなわち、客観的な定量的指標を用いて、特許成立性に関する技術分野の特徴を分析できることを

示したといえる。

5. 考察

以下、「3. 結果」で得られた結果のうち、興味深い結果について考察する。

① 「特許情報等を用いて特許成立性を分析する際、個別の技術分野別に分析した場合の方が、モデルの説明力が良い」ことについて

「1. 序論」で述べたとおり、これまで特許情報を用いて特許成立性を分析した研究はあったものの、それを技術分野別に比較した研究はなかった。今回の「個別の技術分野別に分析したほうが良い」とする結果は、今後の「特許情報を用いた特許成立性の分析」に対する、一つの提言といえる。

② 「機械分野やIT分野においては『請求項文字数』の寄与度が大きい一方、ライフサイエンス分野においては『特許庁応答回数』の寄与度が大きい」ことについて

「請求項文字数」の増加が特許成立性を高めることについては、請求項の文字数の増加により修飾的な記載事項が増えることで発明対象が限定され、厳密性が高まったためと考えられる。そして特に、機械分野やIT分野は、特許成立性を高めるうえで請求項の記載が特に重要な技術分野であるといえる。

一方、「特許庁応答回数」の増加が特許成立性を高めることについては「補正書」や「意見書」の提出により、審査官と出願人の発明に対する認識の差が埋まったためであると考えられる。それは、例えば、審査官の「権利取得のためには、このような発明であるべきという認識」や、出願人の「先行技術との差が、ここに

あるという認識」など等が挙げられる。そして特に、ライフサイエンス分野は、特許成立性を高めるうえで特許庁への応答が特に重要な技術分野であるといえる。

このように特許成立性に影響を与える要因及びその寄与度から技術分野の特徴を分析できるといえる。

③ 「全般的に、引用特許文献数が増えると特許成立性が高まる」ことについて

引用特許文献数の増加により、特許成立性が高まることは一見意外な結果である。

しかし、今回の結果を踏まえると、引用文献数が増えると、本願の技術範囲との差がより明らかになるため、その差についてどのように主張すれば良いかが明確になり、むしろ特許成立性が高まったと考えられる。

④ 「ニューバイオ分野では、出願人が大学等であると特許成立性が高まる」ことについて

このことから、大学等の非営利企業はニューバイオ分野を得意としていることがわかる。

また、医薬化粧品分野では5%水準で有意差は認められなかった(20%の有意水準では認められた)ものの、逆に出願人が「企業」等であると特許成立性が高まる傾向がみられた。この結果から、同じライフサイエンスの分野でも大学等と企業等では、発明の得意分野があると推測できる。今後、出願人による得意分野をさらに分析・検証していきたい。

6. 今後の展望

本研究では、技術分野別の特許成立性に影響を与える要因及びその寄与度を明らかにすることを目的として、客観的な定量的指標のみに基づいたロジスティック回帰分析を行った。この結果、技術分野別に特許成立性を与える要因及び寄与度の違いを明らかにした。

更に、このロジスティック回帰分析による手法を拡張することで、技術分野別に、「客観的な定量的指標のみに基づいて特許成立性を予測することは可能か」という命題が設定される。しかし、この場合、採用する説明変数に対して、より注意が必要である。それは、予測を行う以上、特許成立性の見込みから影響を受けるといえるような説明変数を採用するのは不適當であるといえるからである。

具体的には、今回採用した「補正書提出回数」と「意見書提出回数」の説明変数に注意しなければならない。これらは、特許庁との関係性のみに基づいて値が決定されるという点において、他の説明変数と性格を異にする。すなわち、「補正書提出回数」と「意見書提出回数」は、出願人は、特許庁側の見解を判断した上で、補正書・意見書を提出するため、「出願人が特許庁からの拒絶理由通知書に対し、見込みがあるものを中心にして補正書・意見書を提出して応答している」という可能性があるのである。もしそうである場合は、「補正書・意見書提出回数が特許成立性に影響を与えている」のではなく、「特許成立性(の見込み)が補正書・意見書提出回数に影響を与えている」ことになり、因果関係が逆転してしまっていることになる。このような場合、「補正書提出回数」と「意見書提出回数」を説明変数として採用し特許成立性の予測を行うことは適當ではない。

このため、「補正書提出回数」と「意見書提出回数」を説明変数として除外した場合において、特許成立性に関して予測が可能かどうかを考察するために、「補正書提出回数」と「意見書提出回数」を説明変数として除外したロジスティック回帰モデルを構築し検証を行った(data not shown)。この場合、機械分野、IT分野、医療機器分野において、特許成立性の正答率は70%を超え、ニューバイオ分野、オールドバイオ分野でも60%後半という高い正答率を示した。すなわち、「請求項記載様式」や「先行技術文献数」、「出願人属性」等から抽出した、特許庁の関係性無しで決まりうる定量的指標のみに基づいても、特許成立性の予測が可能であるということが示唆された。

今後は、他出願時期への拡張・他説明変数の追加採用による予測精度の向上、アウトオブサンプルのデータ(ロジスティック回帰分析で使用したデータ以外のデータ)を用いたモデルの頑健性の検証、他技術分野への拡張等による更なる詳細な分析により、特許成立性の予測モデルへの発展性を模索していきたい。

注

- (1)財団法人 知的財産研究所 「特許の経営・経済分析」, 2007, はしがき参照
- (2)安彦元(あびこげん), 中川秀敏「特許明細書における定量的性質の抽出方法に関する研究」知財マネジメント研究 vol.4 (2006)

- (3)永田健太郎, 渡部俊也「日本特許の質に関する実証分析」(日本知財学会 第六回年次学術研究発表会要旨集, 2008, p326-329)
- (4)金井一幸, 上條肇, 清水初志「オープンリソースを用いた新技術の特許化動向分析に関する一考察」Patent vol. 61 No. 6 (2008)
- (5)独立変数間に強い相関がある場合, 回帰係数の推定が不安定になり, 推定値は信用できないとされ, これを多重共線性の問題と言う。太郎丸博「人文・社会科学のためのカテゴリカルデータ解析入門」ナカニシヤ出版, 2006, p116によれば, 相関係数が0.7以上では, いくつかの変数のうちいずれかを除外する必要があるが, 0.7よりも小さくても多重共線性が成り立つ場合があるので注意が必要とされている。
- (6)ただし欠損割合は全体で約1%にすぎない。
- (7)モデル χ^2 検定は, 各説明変数の検定結果(ワルド検定)に優先されるため, 同一の推計モデルで技術分野を比較できるよう, 有意でない説明変数も強制投入した。
- (8)Hosmer&lemeshowの検定結果とは, 回帰式の適合性で実測値と予測値を比較する検定である。重回帰分析で言えば, 重相関係数Rや決定係数 R^2 のようなものである。
- (9)予測精度を表す統計量で, 対馬栄輝「医療系多変量データ解析」東京出版, 2008, p129によれば, 明確な基準はないが, 高い方がよいとある。
- (10)先の「2.2 推計モデル」で述べたように「請求項平均文字数 $\times 0.01$ 」と置き換えても同様の結果である。以下の考察では「【請求項1】文字数 $\times 0.01$ 」と「請求項平均文字数 $\times 0.01$ 」の両方を指すものとして単に「請求項文字数」とする。
- (11)「2.2 推計モデル」で述べたように「意見書提出回数」と置き換えても同様の結果である。以下の考察では「補正書提出回数」と「意見書提出回数」の両方を指すものとして単に「特許庁応答回数」とする。
- (原稿受領 2010. 5. 26)

