

様々な知財関連データの活用

—特許訴訟データを用いた計量分析を例に—

Utilization of various data of Intellectual Property

東京大学政策ビジョン研究センター教授

渡部 俊也

1984年東京工業大学無機材料工学専攻修士課程修了、1994年同大学無機材料工学専攻博士課程修了(工学博士)。民間企業を経て1998年東京大学先端科学技術研究センター情報機能材料客員教授。2001年より同センター教授、2016年から政策ビジョン研究センター教授

日本テラデータ株式会社

胡 章

2014年東京大学農学生命科学研究科応用生命工学専攻修士課程修了、2017年東京大学工学系研究科技術経営戦略学専攻博士課程終了(工学博士)。2017年、日本テラデータ株式会社に入社、ビッグデータ関連の仕事に携わる。

1 はじめに

特許等の書誌情報や審査経過情報を用いて、回帰推計などの統計手法を用いた分析は、イノベーション分野などの学術研究では以前より盛んに行われてきた。さらに特許戦略や研究開発分野の知見を得るための情報ソースとして特許情報を利用することや、特許だけでなく、意匠や商標など幅広い知財情報を利用した解析についても様々な試みが行われている。

例えば筆者らのグループでは、日本企業から東アジア企業への転籍など、発明者の組織間移動について、特許データから情報を抽出した分析(Fujiwara, A. and Watanabe T., 2017.)や、特許以外の知財データとしては、意匠公報を用いた分析なども検討してきている(吉岡(小林)徹・渡部俊也(2016))。最近では、学術目的を超えて、特許の引用情報などでは統計手法を用いた分析を実務的な目的で活用することも一部では試みが始まっている。

この点特許や意匠、商標などの情報は、書誌情報の形式や明細書記載の形式が比較的定式化されていることから、テキストマイニングなどによって情報を抽出しやすく、特徴量を自動で生成することも比較的容易である。

ただこれら以外の知財にかかわる情報の分析を試みようとするとき、定式化されていない様々な文書に関して計量分析を試みることになること、様々な困難が生じる。本稿ではこのような例として、特許訴訟の判決文を用いた分析の試みについて示し、幅広い様々な知財関連データの活用の可能性について議論を行ったものである。

2 特許侵害訴訟の計量分析

特許訴訟の判決文には、勝訴か敗訴か、損害賠償額の多寡、事件の内容や該当する特許番号などが記載されている。さらに審決取消訴訟などでは、特許性の判断なども含まれる。従来、特許の経済的価値の評価は様々な試みがされているが、特許侵害訴訟においてはその価値が、法的有効性に加えて金銭的な評価がなされることから、特許価値の経済的評価に活用できるデータとして期待されることだが、実際は特許と損害賠償事件の判決情報をリンクさせて解析した研究は少ない。例えばLanjouw and Schankerman 2001やChien, 2011, Cremers 2004らは、特許訴訟に巻き込まれやすい特許の特徴を分析し、技術分野や請求項数、後方引用やファミリーサイズなどと有意な関係があることを

示している。特許訴訟に巻き込まれる企業は、偶然の事故では必ずしもなく、自社の特許ポートフォリオ自身にその原因の一端があることを示したものと見える。

一方 Mazzerro et al. 2013 は米国特許訴訟における損害賠償額を被説明変数として、関与する特許の特徴や訴訟における代理人の数などを説明変数として回帰推計を行い75%という高い説明力を有するモデルを構築している。この結果から少なくとも米国特許訴訟における損害賠償額は、ある程度特許と裁判の形式的な情報によって予想可能なものであるということを示唆している。

3 特許訴訟データを用いた特許マネジメントの分析

特許訴訟の分析が限られている背景としては、やはりデータソースとして特許明細書ほど定式化されておらず、記載内容も裁判官の個性に任せられているため、取り扱いが難しいことがあげられる。

筆者らはこのような特許侵害訴訟のデータを用いた特許の価値の推定や、訴訟マネジメントに関するテーマなどにわたる様々な分析を試みてきた。そのいくつかの分析結果を以下に示す

3.1 無効訴訟における有効性判断を正解データとした回帰推計

特許の有効無効の判定について、審決取消訴訟や侵害裁判における最終処分を特許有効性の被説明変数として、特許明細書も書誌情報や審査経過情報から抽出した特徴量を説明変数として、ロジスティック回帰推計（特許が有効か否か）を行った（K. Nagata, et. al. 2008）。説明変数を Table 1 に示す。優先権主張数や請求項文

Table 1 Independent parameters

Parameters	Definition
Domestic_P	Number of Domestic Priorities
Words_AC	Number of words in All Claims (0.1 times)
Words_TC	Number of words in Claim 1 (Top Claim)(0.1 times)
Words_DE	Number of words in the item Detailed Explanation of invention (0.01 times)
Words_DEEBA	Number of words in the rest except Background Art from "Words_DE" (0.01 times)
Effects	Number of words describing "effect" in "Words_DEEBA" (e.g. "may/can" "could" "superior" "useful" "advantageous")
Exa_FPR	Number of Foreign Patent References cited by the examiner
Exa_IFPR	Number of Foreign Patent References added by the examiner
Rejection_N	Number of rejection notices for lack of novelty/inventive step

※ "Issue_W" is coded "1" if validity issues argued contain written description, "0" otherwise. "Issue_N" is coded "1" if a validity issue argued is only novelty or inventive step, "0" otherwise

字数、詳細な説明の文字数、外国文献引用数などに加えて、「〇〇できる」、「〇〇を可能にする」などの効果をあらかず文字数なども加えている。

回帰推計の結果は Table 2 に示した。モデルによっては効果文字などが有意になる傾向があったが、特定の説明変数が大きく効くということはなく、またモデルの当てはまりも必ずしも良くない。

Table 2 Multiple Logistic Regression Analysis

parameters	Model 1	Model 2	Model 3
	Coeff. (S.E)	Coeff. (S.E)	Coeff. (S.E)
Domestic_P	0.2211 (0.142)	0.2232 (0.142)	0.2303 (0.141)
Words_AC	0.0004 (0.001)	0.0004 (0.001)	-
Words_TC	0.0076 (0.006)	0.0076 (0.006)	0.0085 (0.005)
Words_DE	0.0007 (0.012)	0.0002 (0.002)	-
Words_DEEBA	-0.0005 (0.012)	-	-
Effects	0.0062 (0.005)	0.0062 (0.005)	0.0072 (0.004)*
Exa_FPR	-0.2354 (1.244)	-	-
Exa_IFPR	0.5262 (1.261)	0.2917 (0.190)	0.2896 (0.190)
Rejection_N	0.2359 (0.171)	0.2373 (0.170)	0.2411 (0.169)
Issue_W	1.0252 (0.443)*	1.0800 (0.341)**	1.0715 (0.341)**
Issue_N	-0.0680 (0.319)	-	-
E	-0.6630 (0.419)	-0.6721 (0.417)	-0.6814 (0.416)
Constant	-1.9903 (0.375)**	-2.0445 (0.220)**	-2.0453 (0.217)**
N	710	710	710
Nagelkerke R ²	0.075	0.074	0.074

Notes ** and * indicates p value of 1% and 5%, respectively.
Coeff.: Coefficient

Source: K Nagata et.al. 2008

3.2 特許侵害訴訟賠償額（認容額）の回帰推計

特許の有効性ではなく、その特許の金銭的価値に相当すると思われる評価の手がかりとしては、特許侵害裁判において裁判所が認めた損害賠償額の認容額がある。筆者らはこれらの認容額の回帰推計による分析に以前から取り組んできた（西部他 2013）日本における特許侵害訴訟の賠償額の対数を被説明変数として、前方引用数やファミリーサイズ、IPC コードなどを説明変数として、最小二乗法による回帰推計を行った結果を Table 3 に示す。特許の技術的価値を反映していると思われる前方引用数や、特許出願からの経過年数、被告側の代理人数などが有意となるモデルにおいて、調整済み決定係数で27%のモデルを得た。

筆者らは日本における侵害訴訟に加え、米国および

中国でも分析を行っているが、前者で49%、後者で17%のモデルを得た。米国においてはMazzeri et al. 2013の回帰推計モデルでは75%という説明力を得ているが、今回のモデルでも約50%という高い説明力を得たことは興味深い。また有意となる変数は日米中で異なっており、日本においては特許の技術的価値が、米国と中国においてはむしろ特許の経済的価値が認容額に対してより大きな影響を持っていることが示唆されている。ただし米国においては裁判官による裁判においてのみ整合的な結果が得られるが、陪審員の裁判のみを解析した場合、説明力は28%に低下し、有意な説明変数としては代理人の数や懲罰的賠償の有無以外に、特許自身に関係する有意な変数は見出されなかった。

3.3 特許訴訟が企業の特許出願行動に与える影響

特許権者にとって、訴訟自身を不可欠な収益手段として利用する一部のNPEなどを除いて自ら好んで争う企業はいないものと思われるが、自らの特許を実施し事業を行っていく際に、侵害排除のためにやむを得ず原告として訴訟に関与することは一定の確率で生じる。しかし特許権者にとっては訴訟にまつわる様々なリスクがあることも事実であり、侵害裁判で敗訴したり、あるいは勝訴しても十分な賠償額が得られなかったり、さらには裁判にかかる労力や資金などが経営状態にマイナスの影響を与えることもある。特に企業規模が小さく経営資源も乏しい中小企業にとってはその影響は少なくないであろう。

筆者らはこのような観点から、特許権者である中小企業が原告として訴訟に関与した場合の特許出願活動に関する影響について特許訴訟データを用いて解析を試みた(Hu, W., et al. 2017)。被説明変数として訴訟経験後の特許出願数と、特許の質を代理すると思われる特許の前方引用数を用い、特に高額の特許訴訟の場合、どのような影響がどの程度継続するかについて、ラグ変数を説明変数として用いて固定効果モデルによる解析を行った。対象データは過去10年間の116件の中小企業を原告とする特許侵害訴訟である。

Table 4 および Table 5 には、裁判一審終了時及び裁判終審終了時の双方を起点とするラグ変数で特許出願数および特許の前方引用数を回帰した結果を示す。

Table 3 OLS regression results for Log_JP patent infringement awards

	Model 1
Intercept	12.903*** (1.719)
Number of forward citations	0.207*** (0.075)
Family size	0.041 (0.075)
Number of IPC codes	0.061 (0.129)
Age of the litigated patent	0.081** (0.040)
Number of claims	0.009 (0.056)
Number of plaintiff's lawyers	0.001 (0.134)
Number of plaintiff's attorneys	0.206 (0.197)
Number of defendant's lawyers	-0.067 (0.109)
Number of defendant's attorneys	0.781*** (0.203)
Court dummies	Yes
Technological dummies	Yes
Year dummies	Yes
R ²	0.469
Adjusted R ²	0.273
Observations	131

Standard errors in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Source :Hu W et.al. 2016

Table 4. The effect of patent litigation on the maximum number of forward citations of subsequent patent applications

	After the first instance	After the final instance
	Model 1	Model 2
Knowledge Stock	0.002*** (0.0002)	0.002*** (0.0002)
ExtremeCost _{t-1}	0.088 (0.066)	-0.169** (0.078)
ExtremeCost _{t-2}	-0.162** (0.080)	-0.288*** (0.089)
ExtremeCost _{t-3}	-0.460*** (0.108)	-0.522*** (0.113)
ExtremeCost _{t-4}	-0.268*** (0.104)	-0.301*** (0.103)
ExtremeCost _{t-5}	-0.625*** (0.129)	-0.580*** (0.124)
Year Dummies	Yes	Yes
Observations	1158	1158
Log-Likelihood	-2207.2	-2213.7
AIC	4444.4	4457.5

Standard errors in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Source :Hu W et.al. 2017

Table 5. The effect of patent litigation on the maximum number of forward citations of subsequent forward citations

	After the first instance	After the final instance
	Model 1	Model 2
Granted Patents	0.019*** (0.001)	0.021*** (0.001)
Litigation _{t-1}	0.154*** (0.029)	0.213*** (0.030)
Litigation _{t-2}	0.219*** (0.030)	0.192*** (0.032)
Litigation _{t-3}	0.284*** (0.031)	0.354*** (0.031)
Litigation _{t-4}	0.132*** (0.036)	0.126*** (0.037)
Litigation _{t-5}	0.053 (0.041)	-0.180*** (0.046)
Year Dummies	Yes	Yes
Observations	534	534
Log-Likelihood	-4145.4	-4098.9
AIC	8320.8	8227.8

Standard errors in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Source :Hu W et.al. 2017

Table 4 については高額訴訟に関する推計であるが、表から明らかなように5年程度は明らかに出願数の減少が認められる。この傾向は高額訴訟に限らず認められるが、高額訴訟により顕著に認められる傾向である。

一方前方引用数を回帰した Table 5 については、訴訟後4年程度は有意に前方引用数の増加が認められる。このことから中小企業は訴訟を経験することによって、特許のマネジメントに明らかな変化が現れ、特許の質を向上させる一方特許数は厳選していることが示唆される。おそらくは特許訴訟を経験すると、自ら従来実施してきた特許の書き方などの品質が十分でなかったことに気づき、また訴訟費用などの影響からも品質の低い特許を減らし、品質の高い特許を出願するようになるものと思われる。しかしこのような望ましい学習効果は4年程度しか継続しないことも明らかとなった。

4 実務的な応用可能性

本稿で述べた特許訴訟判決文に関する分析は、様々な実務への応用可能性があるものと考えている。特許有効性の予測に関しては、筆者らは3.1に示した研究のうち、訴訟での最終処分ではなく、特許庁の審査結果を用いて同様のロジスティック回帰モデルを用いて、有効性の予測モデルを構築したところ、AUC¹で0.6を超える結果となり、ある程度の予測が可能であることを示している(S.Hido et.al.)。日本においては特許訴訟が少なく、このことで予測モデルとしては十分機能していなかったが、サンプル数が多くなるにしたがって予測モデルの構築も可能になるものと思われる。一方特許有効性に比べて特許侵害訴訟における損害賠償額の回帰推計モデルは、決定係数も高いことが特徴的である。こちらの解析について予測モデルとしての検証を行っているわけではないが、上記の経験からして十分予測モデルの構築も可能ではないかと思われる。すなわち裁判を行う事前に賠償額を予測するというような点での実用性も十分期待できるものと期待される。このような訴訟リスクマネジメントとしての知見に加えて、3.3で述べた企業の研究開

発や特許マネジメントに対する学習効果が存在することも実務的に活用することができる。そもそも訴訟を経験しなくても獲得しなくてはならない特許の質向上を実現するためのマネジメントがあるものと思われる。このようなマネジメントを身に着けるための学習を行い、その効果を訴訟の際の効果と比較してフィードバックするなどのプロセスを構築することも可能ではないかと思われる。

先述したように特許訴訟の判決文は定式化がされておらず、テキストマイニングなどを用いてもなかなか良好なデータセットを構築することがむづかしい。本稿におけるデータセットもかなりの部分手作業で抽出している点は、実務に応用する際には問題になる。一方今後人工知能の応用技術の進歩によって、容易に裁判データから有用なデータセットが構築できれば、その応用可能性は一層広がることが期待できる。

1 Area Under the Curve, ROC 曲線の曲線より下の部分の面積を指す。AUC は 0 から 1 までの値をとり、値が 1 に近いほど判別能が高い。判別能がランダムであるとき、AUC = 0.5 となる。

参考文献

- Fujiwara, A. and Watanabe T., 2017. Knowledge management using external knowledge. *International Journal of Innovation Management* Vol. 21, No. 4, 1750031-1 16.
- 吉岡 (小林) 徹・渡部俊也 (2016)「登録意匠の価値を表す指標—意匠の被引用数についての探索的研究—」『日本知財学会誌』12巻3号: 72頁-95頁 (2016)
- K. Nagata, M. Shima, N. Ono, T. Kuboyama and T. Watanabe, 2008 "Empirical Analysis of Japan Patent Quality", International Association of Management of Technology (IAMOT), CD Proceedings, Dubai International Convention and Exhibition Centre on April 6th - 10th.
- S.Hido, S.Suzuki, R.Nishiyama, T.Imamichi, R.Takahashi, T.Nasukawa, T.Ide, Y.Kanehira, R.Yohda, T. Ueno, A.Tajima and T.Watanabe, 2012 "Modeling Patent Quality: A System for Large-scale Patentability Analysis using Text Mining", *Journal of Information Processing*, Vol.20, No.3.
- 西部壮紀、渡部俊也「特許権侵害訴訟人認容額を用いた特許権価値評価の試み」日本知財学会年次学術大会, 青山学院大学, 11月30日, 12月1日 (2013)
- Hu, W. and Watanabe, T., Nov.18 2016. Patent infringement awards in Japan: an effective legal remedy for patent holders? Asia Pacific Innovation Conference, Kyushu University.
- Hu, W., Yoshioka-Kobayashi T., and Watanabe T., 2017 Impact of patent infringement litigation on the subsequent patenting behavior of plaintiffs from small and medium enterprises, *International Review of Law and Economics*, Volume 51, August, 23-28
- Lanjouw, J. O., & Schankerman, M. 2001. Characteristics of patent litigation: a window on competition. *RAND journal of economics*, 129-151.
- Cremers, K. 2004. Determinants of patent litigation in Germany.
- Chien, C. V. 2011. Predicting patent litigation. *Tex. L. Rev.*, 90, 283.
- Mazzeo, M. J., Hillel, J., & Zyontz, S. 2013. Explaining the "unpredictable" : An empirical analysis of US patent damage awards. *International Review of Law and Economics*, 35, 58-72.
- Li, L. M. 2015. Research on the subject nature and industrial characteristics of statutory compensation in patent infringement. *Modern Law Science*, 37(4), 170-183

