知財訴訟判例文書からの判例統計情報抽出と知財訴訟分析への応用

Extracting judicial information from legal proceedings and analyzing IP lawsuits

野中　尋史: Hirofumi Nonaka [1,2], 酒井　浩之: Hiroyuki Sakai [3], 増山　繁: Shigeru Masuyama [3]

1 豊橋技術科学大学大学院電子・情報工学専攻: Department of Electric and Information Engineering, Toyohashi University of Technology

2 豊橋技術科学大学産学連携推進本部: Head office for industry-academia collaboration, Toyohashi University of Technology

3 豊橋技術科学大学大学院情報・知能工学専攻: Department of Computer Science and Engineering, Toyohashi University of Technology

連絡先住所：愛知県豊橋市天伯町雲雀ヶ丘 1-1 E-mail:nonaka@smlab.tutkie.tut.ac.jp

概要

判例統計情報とは,判例文中から，原告勝訴の割合や損害賠償額をはじめとする必要事項を抽出，データベース化し，それらの統計を取ったものを意味する．特許をはじめとして知財に関わる法学の研究，あるいは，実務において，法律別の訴訟数や訴訟における原告勝訴の割合をはじめとするこれら判例統計情報は，法運用の動向を調べる際に有効なものとなる．しかしながら，このような統計情報，特に条文別など詳細な分類での統計情報を人手により作成する場合は，膨大な労力が必要となる．本研究では，かかる課題を解決すべく，知財訴訟分野を対象に，機械学習により，判例文書を権利種（特許権，著作権などの知的財産権別）などに基づき分類した上で，原告勝訴の確率などの有益な情報を、判例文書の文法構造を利用して抽出するアルゴリズムを開発した．本論文では，提案アルゴリズムの詳細，および，その性能評価，さらには，知財訴訟判例データに適用した場合の簡単な考察を行い，その有用性について示す.

1.背景

知財学の研究あるいは特許業務をはじめとする知財実務において，法律別の訴訟数や訴訟における原告勝訴の割合をはじめとする判例統計情報は，法運用の動向を調べる際に有効なものとなる．例えば，文献[1]は，特許法 102 条 1 項に基づく損害賠償額算定に関して考察しているが，この際，同法 102 条 1 項-3 項など損害賠償額を規定する条文ごとの損害賠償認定率について文献[2]の統計を用いて比較し， 同法 102 条 1 項の有効性を明らかにしている．また，法実務においても，例えば，知財訴訟の場合，どの程度，権利者である原告が勝訴しているか，判例統計を利用することにより知ることができ，訴訟時の戦略決定に役立つ.

このような判例統計情報の利用は，上記に示したように単にデータをそのまま使用して統計的傾向を調べることに留まらない．統計モデル構築に利用すれば，法運用メカ

ニズムの解明や，経営学との融合による知財法が経営にあたえる影響分析などができる．例えば，文献[3]は，実証的な判例研究についての現状・展望を示したものである．この中で，判例等から抽出した情報についてロジスティック回帰モデルなどの統計学的手法による解析の必要性が記述してあり，判例統計情報の整備のニーズは高い．また，判例統計情報と経営・経済データとの相関を取ることにより，訴訟が及ぼす経済効果も導出できるようになる．例えば，知財訴訟，特に特許民事訴訟に関し，判例統計情報(データベース)を用いて，勝訴，ないし，敗訴した企業を特定し，その訴訟の事前・事後の経営指標（例えば株価）の変化が統計的に有意なものであるか否かを計測する金融工学的手法の一種であるイベントスタディ法を用いることで，判例が経営に影響を与えたかどうか，また，どの程度与えたのかを推し量ることができる．このように，判例統計情報の整備が重要であるが，継続的，体系的，かつ，詳細に整備された統計情報・データベースはない．文献[2]に記載された詳細な判例統計情報なども単発的に収集されたものであり，継続的にとられているものではない．これは，膨大な判例について，人手を使ってデータベースを作成する際，多大な労力を要するためと考えられる．そのため，情報工学的な見地より，判例統計情報を整備する必要がある．しかしながら，現在，判例（裁判文書）の概念検索[4]などのシステムが開発されているが，原告勝訴の確率や損害賠償認定率などの統計情報を自動取得する手法の研究はなされてない．

2. 提案手法

本研究では，上記，課題を解決する手法を提案する．特に問題となるのは，民事行政訴訟別，権利種別などの分類や，訴訟における勝訴敗訴の判定である．ここで，民事訴訟（図 1）とは私人間の生活関係に関する紛争（権利義務に関する争い）につき、私法（ここでは，特許法等の知財に関連する法）を適用して解決するための訴訟を指す．一方，行政訴訟（図 2）とは行政事件に関する訴訟であり，特許法をはじめとする知財に関連する法においては，特許庁の審決をめぐる取り消しなどが該当する．

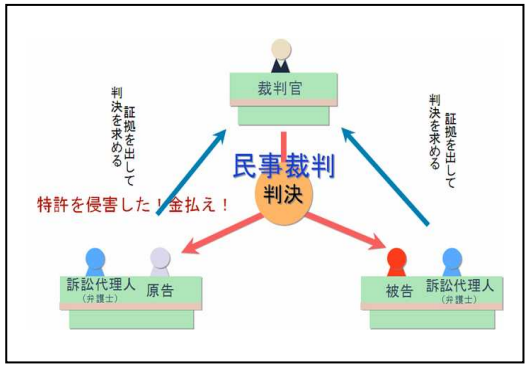


図 1. 知財民事訴訟

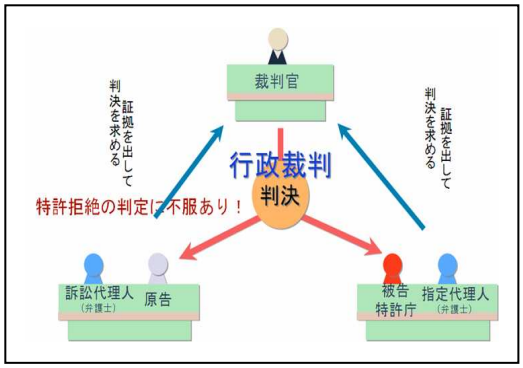
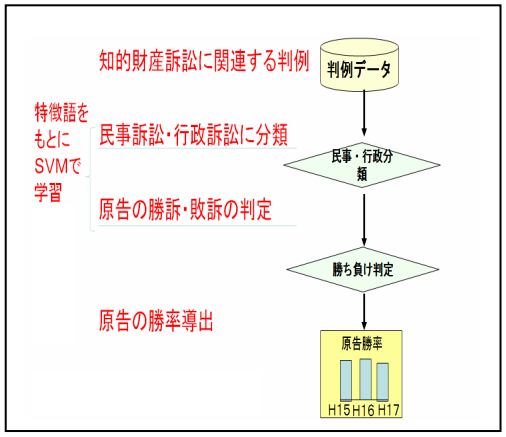


図 2. 知財行政訴訟

これらの情報は判例中において，明示的にタグ付けされているわけではなく，判例文書中の主文などを解析することにより，はじめて得られる情報である．そのため，本研究では，機械学習アルゴリズムを用いた分類・判定手法を開発した．具体的には，判例文書を形態素解析（各文書を単語単位に分割すること）した上で，判例文書の構造中，適切な部分を切り出し，その部分の特徴語のみを素性（機械学習において学習する際の特徴量）として後述のサポートベクターマシン(Support vector machine；以降 SVM と表記)[5]による学習を行い，学習により作成した分類器により，民事行政訴訟別，権利種別などの分類や，訴訟における勝訴敗訴の判定を機械的に行う．全体の手法の概要を図 3 に示す．



」

図 3. 手法概略

なお，以下の 2 項目の有益な情報は，機械学習を用いずとも抽出可能である．まず，損害賠償額については，主文中（裁判所が下した），および，請求中にある金額表現の抽出により特定できる．さらに，主文中の損害賠償額（裁判所が認定した損害賠償額）で請求中の損害賠償額（原告が請求した損害賠償額）を除すことで，損害賠償認定率も計算できる．また，技術分野の特定においては，特許番号をフォーマット（特許第〇〇〇〇〇〇号など）に基づいて抽出し，別途特許データベースと併用することで得られる IPC 分類により行うことができる．以下，本研究で提案する手法，特に機械学習を行う部分について詳しく述べる．

2.1. SVM

本提案手法では，機械学習アルゴリズムとしてサポートベクトルマシン（SVM）を使用する．SVM は， 教師あり学習を用いる識別手法の一つであり，汎化能力が高い優秀な学習モデルの 1 つであることが知られている．本研究では，線形カーネルを使用した．

2.2. 判例文書の構造中，適切な部分の選択

機械学習を行う上で，その素性は特徴を判別するものでなくてはならない．このため，全文を対象にするよりも，分類を行うのに，よりふさわしい部分のみを選択する方が望ましい．そこで，分類の特徴に併せて適切な部分の選択を行う．民事行政訴訟別の分類を行う際は，主文，請求，事案の概要部の要旨を用いる．これは，民事訴訟の場合，主文や請求，事案の概要部の要旨には，「被告は〇〇を払え」などの文言が多く登場し，行政訴訟の場合は，「特許庁が下した審決を取り消す．」などの文言が出現しやすいなどの統計的偏りがあるため

である．また，権利種別の分類においても同様の部分を使用する．これは，上記と同様に権利種の特徴を現す語が多く出現するためである．例えば，商標の場合，「標章」など商標固有の表現が多く出現する．一方，判例において，原告が勝訴（一部認容も含む）したかどうかを判定するには，主文を用いればよい．

2.3.素性選択

本研究では，新聞記事からの交通事故事例解析で，酒井ら[7],[8]が用いた手法により「事案の概要」部中の特徴語を抽出し，それを素性として分類を行う．具体的には，以下のステップで行う．

Step 1:2 値（一方を正例と定義し，と一方を負例と定義する．）にわけて教師データを用意する

Step 2:式（２）より構成される重み $Wp(t_i, S_p)$の計算（式（１））を行う．同様に負例の重み $W_n(t_i, S_p)$も計算し，その２倍超になる語を特徴語とする．以下，式の説明を行う．

$$W_p(t_i, S_p) = P(t_i, S_p)H(t_i, S_p) \quad \cdots(1)$$

$$P(t_i, S_p) = \frac{Tf(t_i, S_p)}{\sum_{t \in Ts_{S_p}} Tf(t, S_p)} \quad \cdots(2)$$

ただし，

$P(t_i, S_p)$: 正例の文書集合S_p における語t_i の出現確率

S_p: 訓練データにおいて正例に属する文書集合

$Tf(t_i, S_p)$: 正例の文書集合S_p に含まれる語t_i の数

T_{sSp}: 正例の文書集合S_p に含まれる語の集合

$H(t_i, S_p)$: 正例の文書集合S_p に含まれる各文書における語t_i の出現確率に基づくエントロピー

ここで，エントロピーは統計的な偏りを表す量であり，偏りがなく満遍なく出現する語ほど大きな値となる．このため，Step 2 を満たすものは，正例集合に満遍なく出現し，かつ，出現頻度も大きいものであり，逆に,負例集合には，少数の文章にしか出現しないものとなり，正例に関する特徴語といえる．上記により，正例に偏って出現する特徴語を素性として，効果的・効率的な分類を行うことができる．

3.評価実験

評価実験に使用する判例データは，LEPOSYSTEMS 社の平成 19 年度六法全書判例データ（知財訴訟約 4，000 例収録）を用いた．以下，各項目の評価と考察を示す．ページ数制限の都合上，ここでは，民事行政訴訟別の分類，権利種別の分類，原告勝訴の判定の性能のみを示す．

3.1.民事行政訴訟別の分類

H17 年度の知財訴訟判例から，正例（民事訴訟判例）204 個，負例（行政訴訟判例）204 個をを教師データとして選択し，それぞれの権利種別に SVM による識別器を作成

した．この識別器により，H18 知財訴訟判例 606 件を分類したときの性能を表 2 に示す．なお，検索システムの評価に使用される再現率，適合率に加えて，全体の判別の正確性も重要となるため，正確度も指標とした．

表 2.分類性能

正確度(全体の精度)	0.97
適合率（正例の精度）	0.98
再現率（正例の網羅性）	0.92

結果より，かなり精度よく分類できていることがわかる．これは，行政訴訟，民事訴訟では，主文などの文言に偏りがあり，それをうまく生かし，特徴を判別できた結果だと考える．

3.2.権利種別の分類

H19 年度，H17 年度の知財訴訟判例から，特許権は正例 45 個，負例（特許権以外の権利種に関する判例）45 個を，それ以外の権利種に関しては正例 15 個，負例 15 個を教師データとして選択し，それぞれの権利種別に SVM による識別器を作成した．この識別器により，H18 年度知的財産民事訴訟 204 件の分類を行うことで，評価を行った．特許権に関して，教師データが多いのは，裁判で争われる内容の多様性が大きい（条文数が多い）ためである．適合率，再現率の結果は，表 3 の通りとなった．

表 3．分類性能

	特許権	商標権	意匠権
再現率	0.8	0.91	1
適合率	1	1	1
	不正競争防止法		著作権
再現率	1		0.93
適合率	0.96		1
	実用新案		
再現率	1		
適合率	1		

いずれも適合率は高く，本手法アルゴリズムで抽出した例の 9 割から 10 割が正答となっている．また，再現率も高めである．これは，各権利種別に特徴的な専門（商標だと「標章」など）が使用される傾向が高いため，分類しやすかったと推測する．

4.3.原告勝訴の判定

H17 年度の知財訴訟判例から，正例（原告勝訴）59 個，負例（原告敗訴）59 個を教師データとして選択し，それぞれの権利種別に SVM による識別器を作成した．この識別器により，H18 知財訴訟判民事訴訟例 204 件を分類したときの性能を表 3 に示す．なお，検索システムの評価に使用される再現率，適合率に加えて，全体の判別の正確性も重要となるため，正確度も指標とした．

表 4.分類性能

正確度(全体の精度)	0.94
適合率（正例の精度）	0.85
再現率（正例の網羅性）	0.97

結果より，精度よく分類できていることがわかる．

4. 知財訴訟判例データの解析に応用した例の考察

上記より，知財訴訟判例より，非常に高い精度で原告の勝率などが抽出できることを示した．ここでは，上記で評価した識別器を使用して知的財産民事訴訟における原告勝訴の判例を抜き出し，データベースを構築し，年度ごとの原告の勝率導出に応用した例を用いて，知財司法運用システムの考察を行う．以下，結果を図 4 に示す.

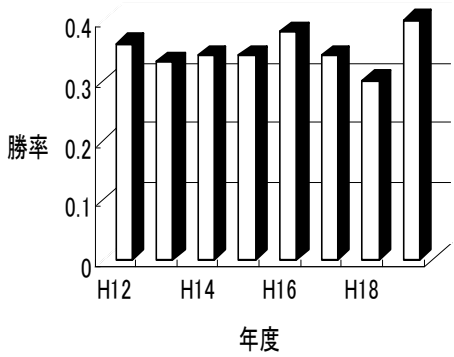


図 4.知的財産民事訴訟原告勝率

図 4 より，度による差異はほとんどなく，原告の勝率が 3 割程度で推移していることが分かる．すなわち，年度毎に極端に原告の勝率が変化することなく，安定した法運用がなされている.

また，訴訟になった場合，権利範囲の認定が厳しく，原告の勝率は 5 割に満たず，3 割程度であることも示している．このことは，特許権をはじめ，知的財産権は権利取得・維持に多大なコストを要するものの，権利取得により無条件に権利侵害を防ぐものとはならないことを示唆している．よって，知財訴訟を通じて，権利侵害を防ぐ強い知財権を確立するためには，出願時の権利範囲の検討を十分に行うことが非常に重要であることが分かる.

5. 当日の発表について

当日は，上記開発した，情報抽出技術を利用して作成したデータベースを利用し，以下のような研究についての発表も行う.

5.1. 判例統計を用いた詳細な知財訴訟分析

本提案手法により，技術分類や損害賠償額などの別に判例統計を作成することができる．よって，これを応用して，技術分類毎の原告の訴訟における勝率や根拠条文別などの状況別による損害賠償認定率・額などを導出することも可能である．今後は，このような詳細に分類された判例統計を使用して，実証的な法学研究を行っていく.

5.2 知財訴訟判決を決定する要因の分析・予測モデルの開発

知財訴訟において，その判決に影響している因子としては，特許文書の質（たとえば，実施例が多いことや対象となる特許が複数存在し，広い権利範囲であれば原告である権利者に有利となる）や代理人の別（特に経験豊富な代理人であれば訴訟に勝つ可能性が高いかもしれない）などを考えることができる．そこで，判例テキスト文や特許文書中から，実施例の占める割合や請求項の数，請求項に使用される語，代理人の訴訟経験数などを抜き出し，これらを素性として，決定木，または，ロジスティック回帰モデルなどによる解析を行い，どのような素性（例えば請求項の数など）が判決に影響を及ぼしたのかの分析を行う.

具体的には，本提案手法により，原告が勝訴した場合，および，敗訴した場合に判例を分類した上で，原告が勝訴した場合，敗訴した場合に分けて，特許文書データより特許の質に関するデータを導出，また，代理人の別などを各判例より抽出し，決定木などのモデルにより分析することが考えられる．さらに，構築したモデルを利用すれば，判決予測に役立てることができる．

5.3.知財訴訟判決が経営に与える影響の分析

計量ファイナンスの分野においては，後述する「イベントスタディ法（ケーススタディ法)」とよばれる手法がある．これは，あるイベントが株価（などの経営指標）に影響をあたえたかどうかを統計的な検定を使って明らかにするものである．例えば，文献[9]は，イベントスタディ法を利用し，新薬開発が承認された企業について，そのニュースが株価に及ぼす影響を調べている．また，文献[10]は，企業のリストラが株価へ及ぼす影響をイベントスタディ法により実証している．

一方，知財法の分野では，判例が経営・経済に及ぼす効果を測定することなど，従来の法学では，要求されてこなかった事項も研究の対象となっている[1]．これは，経済活性化を目的とする知的財産法が，真にその目的を達しているのか否かを調べることのニーズが大きいためである．

現在，知財法と経済学の分野の研究は，理論経済学に基づくものが主流である．例えば，文献[1]では，特許法による過度の規制により生じるコストと侵害により生じるコストのトレードオフ関係を示し，最適な特許法運用の水準があることを示唆している．ただし，この研究は実データを用いたものではなく，具体的な最適水準を示すものではない．このように，現状の知財法と経済学の研究では，理論的示唆が中心となっている．

そこで，上記，イベントスタディ法，および，これまでの研究で得た，民事行政訴訟種別・権利種別・原告勝訴敗訴の別により分類したデータを用いて知財民事訴訟の判例が株価に影響を及ぼすかどうか実証する．これにより，原告勝訴の判例が株価に影響を与えるという結果がでれば，法運用が権利者側にたったものであることがわかり，また，権利者は自身の知財権を効果的に使い，株主の利益につなげていることが明らかになる．さらに，株価予測の研究に発展させていくことも可能である．一方，影響がないとすれば，権利者は，株主利益最大化の点で，知財権を最適に運用しているかどうか疑義があることがわかる．また，法運用は，侵害者寄りであり，損害賠償額認容の程度が低いことを示唆することになる．

6.まとめ

本研究では，機械学習アルゴリズムを用いた分類・判定手法を開発した．具体的には，判例文書を形態素解析（各文書を単語単位に分割すること）した上で，判例文書の構造中，適切な部分を切り出し，その部分の特徴語のみを素性（機械学習において学習する際の特徴量）として後述の SVM による学習を行い，学習により作成した分類器により，民事行政訴訟別，権利種別など

の分類や，訴訟における勝訴敗訴の判定を自動的に行った．その結果，それぞれの実験において，高い再現率と精度を示し，実用的に本提案手法が使用できることを示した．

また，実際に本提案手法を使用して，知的財産民事訴訟における原告勝訴の判例を抜き出し，データベースを構築し，年度ごとの原告の勝率を導出し，知財司法運用システムの考察を行った．その結果，原告の勝率は 3 割程度と低い値であることがわかり，権利取得により無条件に権利侵害を防ぐものとはならないこと，知財訴訟を通じて，権利侵害を防ぐ強い知財権を確立するためには，出願時の権利範囲の検討を十分と行うことが非常に重要であることを示唆した．

今後は，判例統計を用いた詳細な訴訟分析，知財訴訟判決を決定する要因の分析・予測モデルの開発，知財訴訟判決が経営に与える影響の分析などに本手法を応用していく．

謝辞

本研究は，電気通信普及財団研究調査助成，および，文部科学省グローバル COE プログラム「インテリジェントセンシングのフロンティア」の支援により行われた．

参考文献

［1］宮原 耕史，「特許法 102 条第 1 項による損害賠償額の算定についての考察」，日本知財学会誌,第 3 巻第 3 号,(2007)

［2］特許庁資料，「工業所有権審議会損害賠償等小委員会報告書」(1998)

［3］湯淺墾道，「実証的判例分析の現状と課題」，情報ネットワーク法学会第 3 回研究大会，(2003)

［4］江越裕紀，片上大輔，新田克己，「判例の構造を利用した判例文書検索」，情報処理学会第 48 回デジタルドキュメント研究会，(2005)．

［5］ C. Cortes and V. Vapnik, Support-vector networks, Mach. Learn. 20 (3), pp. 273-297, (1995)

［6］ Nello Cristianini, John Shawe-Taylor 著，大北剛訳，「サポートベクターマシン
入門」，共立出版，2005 年

［7］酒井 浩之，梅村 祥之，増山 繁，「交通事故事例に含まれる事故原因表現の新聞記事からの抽出」，自然言語処理，vol.13, no.4, 99 項，(2007)

［8］Hiroyuki Sakai, Shigeru Masuyama, “Cause Information Extraction from Financial Articles Concerning Business Performance”, IEICE Trans. Information and Systems, Vol.ED.,No.4, pp.959-968, (2008)

［9］鈴木公明，「新薬関連イベントに対する株価反応に関する実証研究」，日本知財学会第六回年次学術研究発表会要旨集 334-337 項，(2008)

［10］清水一,山崎尚志「人員削減と株価パフォーマンス」,神戸大学経営学研究科ディスカッションペーパー2007-14，(2007)

［11］MacKinlay, A. C. , “Event studies in Economics and Finance,Journal of Economic Literature”,Vol.35, Issue 1(March 1997) pp13-39, (1997)

［12］Kothari, J., A. Shelifer, and R.W. Vishny,”Measuring Long-Horizon Security Price Performance”, Journal of Financial Economics 43(3),pp30